Towards gene therapy for cystic fibrosis: Enhanced Green Fluorescent Protein as a reporter of promoter activity

Wendilywn E Walker

Ph.D. Thesis The University of Edinburgh 2002

I Declare:

(a) That this thesis has been composed by myself
(b) That the work is my own, unless otherwise stated
(c) That the work has not been submitted for any other degree or professional qualification

Wendilywn E Walker
May 25, 2002

...for my father, Christopher.

Acknowledgements

Thanks to Chris Boyd for taking me on as a Ph.D. student, for believing in me when I had doubts and for pushing me to go the extra mile. Thanks for providing structure for my Ph.D. project and for helping me with the experimental, scientific, and literary aspects alike. Thanks also for stimulating conversations (intellectual and otherwise) and for your kindness and support.

Thanks to David Porteous for welcoming me into his lab, for his wisdom and experience as an advisor, and for creating a supportive environment for experimental research in the Medical Genetic Section.

Thanks to Heather Davidson for her devotion as lab manager, for help with the preliminary vector design, for teaching me to use the HGMP programs and numerous experimental techniques, and for helping me at every opportunity in the completion of my experiments.

Thanks to Ann Doherty for her patience in teaching me numerous molecular biology and cell culture techniques, for her collaboration in establishing and characterizing the sheep primary cultures, and for being a great neighbor at the bench.

Thanks to Varrie Oglivie and Cath Payne, for their help troubleshooting my experiments, and for a stimulating discussion during group meetings.

Thanks to Peter Thorpe and Barbara Stevenson, for their help and advice on cell culture, the gathering and analysis of FACS data, and for taking an interest in my work.

Thanks to Xinsheng Nan for his wise advice on many aspects of the project, and for his valuable insight in interpreting experimental results.

The foot and mouth epidemic was a significant hurdle to establishing the sheep primary cultures. Special thanks go to Jock at the Moredun Research facility, who provided the sheep tracheae whenever possible, and to Ann Doherty, who allowed me to use most of the available tracheas at the end of this project, in order to finish my *ex vivo* studies.

Thanks to Donald Davidson for helping me establish mouse primary tracheal cultures prior to the sheep cultures, and for advice on the correct appearance and electrical resistance calculations for these cultures.

Thanks to Hazel Davidson-Smith for her help with the Saint-Mix method and to both Hazel and Laura Hyndman for their help in establishing the sheep primary cultures.

Thanks to Eric Miller, for allowing me to use the Imperial Cancer Research Campaign's FACSCalibur machine for the two FACS sorting experiments, and twice in an emergency, when our FACS machine was broken (or had mysteriously disappeared!), and also for his wise advice on this complicated machine.

Thanks to Steve Mitchell for processing my samples for electron microscopy, and for being exceptionally helpful and accommodating.

Thanks to Angie Fawkes for running thousands of sequencing samples for me on the ABI, for help troubleshooting the sequencing reaction when it wasn't working and for brightening the lab with her cheerful disposition.

Thanks to Naomi Wray for her help with the statistical analysis.

Thanks to Alex and Slava at A2 for their help with software.

Thanks are also due to those who helped me with techniques that widened my scientific knnowledge and experience, even though they did not end up featuring prominently in my research. Thanks to Shirley McBride for demonstrating the lung-slice culture technique. Thanks to Paul Perry for his wisdom and generous help with the digital microscopy equipment.

Thanks to my fellow students, Rachel, Sarah, Simon, Jean, Helen, Dawn and Danny, for listening to my student presentations, for their helpful suggestions, and especially for their camaraderie.

Thanks to everyone in the Medical Genetics Section for their help and support throughout my Ph.D.

Thanks to everyone at the MRC HGU for their help and support during the first year of my Ph.D., and to the staff of the Clinical Genetics Section, and the groups on the MMC second and third floors, for allowing me to share their equipment for sequencing, tissue culture, and x-ray film development.

Thanks to the CVCP for funding in the form of an ORS scholarship.

Thanks to my family - Christopher, Jeannie, Julian, Cameron, Dylan and Kate - for helping me to keep my spirits up.

And last (but certainly not least) thanks to Roger for picking up the loose ends when I had run out of steam, for allowing me to appropriate his PC, and for his love and support: without him this thesis would not have been possible.

Abstract

Cystic Fibrosis (CF) is the most common lethal inherited disease, affecting ~1/2000 live births. Although the genetic lesion, a mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, has been elucidated, the exact mechanism whereby this causes the debilitating disease phenotype is unclear. CF patients are prone to repeated bacterial infection of the lung; leading to fibrosis of this tissue, and eventually respiratory failure. Gene therapy has the potential to cure CF: by introducing a normal copy of the CFTR gene into epithelial cells of the lung, it may be possible to abolish the lung phenotype. While early gene therapy vectors utilised a strong viral promoter (e.g. P_{CMV}) to drive expression of the CFTR cDNA, it has become apparent that this approach creates problems: expression is short lived and may not be targeted to the correct cell types at appropriate levels.

CFTR expression is tightly restricted, both temporally and spatially in affected tissues, notably the lung. Putative regulatory elements have been identified upstream of the CFTR gene, both proximally and distally. In addition, a DNase I hypersensitive site has been identified in intron 1 of the gene, and has been associated with an increase in CFTR expression in cells of the gut. Our approach is to include genomic components of the CFTR locus fused to cDNA in gene therapy vectors. These 'genomic context vectors' (GCVs) may be capable of recapitulating the endogenous CFTR expression pattern.

The aim of this project was to investigate the use of the Enhanced Green Fluorescent Protein (EGFP) as a reporter of CFTR promoter activity. Six vectors were created coupling portions of the CFTR locus to EGFP in GCVs. Small plasmids were made by conventional cloning procedures, while large PAC vectors were made by a double recombination method employing both homologous and Cre recombinase/loxP recombination. Some vectors contained an internal ribosome entry site (IRES) to allow separate translation of proteins from a single mRNA:

Name	Size	Promoter	CFTR Introns	Genes
p1kbcfproEGFP	4.9kb	1kb CFTR 5' region		EGFP
$_{PAC}$65kbcfproEGFP	81kb	65kb CFTR 5' region		EGFP
pCFTRiresEGFP	11kb	P_{CMV}		CFTR and EGFP (with IRES)
$_{PAC}$RC1iresEGFP	86kb	65kb CFTR 5' region		CFTR and EGFP (with IRES)
$_{PAC}$RC2iresEGFP	110kb	65kb CFTR 5' region	Intron 1	CFTR and EGFP (with IRES)
$_{PAC}$RC2cmvEGFP	111kb	P_{CMV}		EGFP
pEGFP-N (Clontech)*	4.7kb	P_{CMV}		EGFP

* The Clontech pEGFP-N vector was used as a control

These vectors were transfected into permanent cell lines COS7, MDCK-$_{IOWA}$, T84 and CaCO2, in order to assess the effects of the genomic context elements upon EGFP expression. Several transfection methods were compared to optimise transfection efficiency, including the liposome DOTAP, the Saint-Mix™ synthetic amphiphile delivery system, a Polyethylenimine (PEI) method, and the LID method. A novel transfection method, the 'SID' method (incorporating Saint-Mix ™ and the integrin targeting peptide P6), showed early promise. Ultimately, the LID method was chosen for further studies, as this method was both efficient and consistent.

The proximal CFTR 5' region in the p1kbcfproEGFP vector drove expression of the EGFP transgene at low levels in every cell line analysed. This is in agreement with previous reports that show basal levels of CFTR expression driven by this proximal 'housekeeping' region. The additional upstream region in the $_{PAC}$65bcfproEGFP vector did not appear to modulate expression in any of the cell lines analysed. A comparison of the twin vectors $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP, which differ only in the absence or presence of CFTR intron 1 respectively, showed similar levels of expression in the COS7 and MDCK-$_{IOWA}$ cell lines. Thus, the intron 1 element does not seem to alter expression in these non-gut cell lines; this is consistent with reports that show regulation of CFTR expression in

response to the intron element to be specific to cells of the gut epithelium. An investigation of the intron 1 element in gut cell lines T84 and CaCO2 was thwarted by low transfection efficiency in these cell lines, coupled with inefficient translation from the IRES.

A comparison of pEGFP-N and $_{PAC}$RC2cmvEGFP revealed that large PAC vectors show an intrinsic reduction in expression in comparison to their small plasmid counterparts. Further experiments showed that this was not an effect of vector copy number, and that the effect could not act in trans upon a co-transfected molecule. These studies also revealed an unexpected interaction: diluting a reporter plasmid with an anonymous plasmid may actually increase its transfection efficiency.

Permanent cell lines form an important starting point for transfection studies, but ultimately have proved to be poor models for *in vivo* gene transfer. An *ex vivo* primary air interface sheep tracheal culture was utilised as a more realistic model. These cultures were characterised by electron microscopy, and demonstrated features of the native tracheal epithelium. A measure of transepithelial resistance confirmed the presence of tight junctions. Cultures were transfected with several of the genomic context vectors. While PAC vectors had shown a dramatic reduction in expression relative to their small plasmid counterparts in the *in vitro* studies, only a small reduction was seen in the *ex vivo* cultures, thus PAC vectors, such as GCVs, may provide a promising approach for gene therapy studies.

Table of Contents

List of Tables **page #**

List of Figures **page #**

List of Figures (continued) **page #**

Abbreviations

AAV	adeno-associated virus
AMV-RT	AMV reverse transcriptase
ASL	airway surface liquid
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BSA	bovine serum albumin
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
Cl-	chloride
CRE	cAMP response element
Cre	Cre recombinase
DHS	DNase I hypersensitive site
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
EGFP	Enhanced Green Fluorescent Protein
EKLF	erythroid-specific krüppel-like factor
ENaC	epithelial sodium channel
EtOH	ethanol
ExoSAP-IT	exogenous shrimp alkaline phosphatase
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
FL1-H	green fluorescence
FL2-H	orange fluorescence
FL3-H	red fluorescence
FSC	forward scatter
GCV	genomic context vector
GFP	Green Fluorescent Protein
H_2O	water
IRES	internal ribosome entry site
NBD	nucleotide binding domain
TM	transmembrane
IPTG	isoproylthiogalactoside
LID	liposome – integrin targeting peptide – DNA complex
LCR	locus control region
mRNA	messenger RNA
^{32}P	phosphorous isotope 32
P_{CMV}	cytomegalvirus immediate early promoter
P. aeruginosa	*Pseudomonas aeruguginosa*
PAC	P1 artificial chromosome
PD	potential difference
PEI	polyethylenimine
PCR	polymerase chain reaction
PI	pancreatic insufficient
PNK	polynucleotide kinase

PKA	protein kinase A
PMA	phorbol myristyrate acetate
PMR	region of pyrimidine/purine assymetry
Ppi	pyrophosphate
PS	pancreatic sufficient
R domain	regulatory domain
RNA	ribonucleic acid
RNaseIn	RNase inhibitor
RPM	revolutions per minute
RT-PCR	reverse transcription-PCR
S. aureus	*Staphylococcus aureus*
SEM	scanning electron microscopy
SDS	sodium dodecyl sulfate
SID	Saint-Mix™ – integrin targeting peptide – DNA complex
SSC	side scatter
TEM	transmission electron microscopy
UV	ultraviolet
wt	wild type
YAC	yeast artificial chromosome
%[f]	percent of cells transfected
X_f	average fluorescence intensity

Units of measurement

U	Units
°	Degrees
C	Celcius

Volume

L	Litres
ml	millilitres
µl	microlitres

Weight

G	grams
mg	milligrams
µg	micrograms
ng	nanograms
pg	picograms
fg	femtograms
ag	attograms

Number

mol	moles (6.02252×10^{23})
mmol	millimoles
µmol	micromoles
nmol	nanomoles
pmol	picomoles
fmol	femtomoles
amol	attomoles

Concentration

M	Molar
mM	millimolar
µM	micromolar

Size

M	metre
cm	centimetre
mm	millimetre
µ	microns (micrometres)

Electricity

Ω	ohms
ms	milliseconds
V	Volts
kV	kilovolts

Chapter 1: Introduction

1.1 Cystic Fibrosis

1.1.1 A Brief History of Cystic Fibrosis

Anderson (Anderson, 1938) first described the disease of Cystic Fibrosis (CF). She reported on a cohort of celiac patients displaying diarrhea and steatorrhea, due to a pancreatic enzyme deficiency; this disease followed a simple recessive pattern of inheritance, suggesting that a single gene was responsible (Anderson and Hodges, 1946). Cystic fibrosis was observed in the pancreatic tissue, and hence the disease was named.

In 1983, Quinton demonstrated a chloride (Cl-) transport deficiency in skin biopsies from CF patients (Quinton, 1983). Further studies revealed that this defect was not specific to the skin, but applied to a range of CF tissues (Lin and Gruenstein, 1987; Chen et al, 1989; Bubien et al, 1990).

Linkage analysis and positional cloning techniques, such as chromosome walking and jumping, indicated that the CF gene was positioned on the long arm of chromosome seven (Tsui et al, 1985; Rommens et al, 1989). Eventually, a candidate gene was identified in this region (Riordan et al, 1989; Kerem et al, 1989; Rommens et al, 1989). Several lines of evidence indicated that this gene was the site of the CF lesion, including the gene's location, predicted protein structure, and localization of mRNA transcript to CF-affected tissues. Most convincingly, a 3-bp deletion (ΔF508) was found in this gene on 70% of CF chromosomes; this mutation was completely absent from normal chromosomes (Kerem et al, 1989). This gene was named the cystic fibrosis transmembrane conductance regulator (CFTR).

1.1.2 CF Patient Profile

CF affects multiple organs. Blockage of the meconium (first feces) and subsequent intestinal obstructions are common, due to poor absorption in the gut. In patients with pancreatic insufficiency mutations (see Section 1.1.5), the exocrine glands of the pancreas are destroyed; hence, pancreatic secretions (containing enzymes important for digestion) are not produced. A high sodium chloride concentration is present in the sweat. The disease also affects the liver and reproductive organs: males

with CF are infertile, while females with CF have reduced fertility (Phillipson, 1998). However, the defining feature of CF is the pathology of the lung: patients are prone to repeated bacterial infection of the lungs, resulting in fibrosis of this tissue, and eventually respiratory failure unless a lung transplant is performed (Davidson and Porteous, 1998; Gaiso, 1998; Korst et al, 1995).

The lung disease in CF follows a generally progressive course, marked by a number of turning points (Pilewsky and Frizzell, 1999). At birth, the lung appears to be normal. The earliest pathology of the CF lung is a mucus obstruction of the small airways, accompanied by inflammation and neutrophil infiltration of the bronchiolar walls (Korst et al, 1995). There is some evidence that an inflammatory response might be present in infants below six months of age: such a response might even precede bacterial infection of the lung (Cantin, 1995), although this hypothesis is controversial.

By two years of age, the bronchioles and bronchus become infected with *Haemophilus influenzae* and *Staphylococcus aureus* (Pilewsky and Frizzell 1999). This is accompanied by further inflammation, and thick, sticky mucus, which obstructs the small airways. Mucociliary clearance is impaired, resulting in a failure to remove bacteria from the lung. Persistent inflammation and bronchiecstasis develop between two and ten years. As CF airway disease progresses, the pathogen profile changes. *Pseudomonas aeruginosa* is an opportunistic pathogen that is commonly present, but rarely causes infection, in healthy individuals; this has been implicated as the major pathogen in CF lung disease. The CF lung first becomes transiently infected with *P. aeruginosa*, but eventually mucoid strains emerge, colonizing the lung. Over the next decade, inflammation and bronchiecstasis are accelerated and progressive airway obstructions result. Eventually, respiratory insufficiency develops, leading to death.

1.1.3 Treatments

The use of pancreatic enzymes and nutritional supplements is highly effective in treating the CF pancreatic insufficiency, leaving lung disease as the major cause of mortality (accounting for ~95% of deaths among CF patients). Aggressive

physiotherapy and antibiotic regimes have dramatically increased the lifespan of CF patients (Korst et al, 1995; Gaiso, 1998). Recent advances in CF therapies include: Pulmozyme®, a drug that reduces the viscosity of the mucus (by digesting DNA remnants of lysed neutrophils); ibuprofen, which has anti-inflammatory effects; and TOBI™, a tobramycin solution for inhalation (CF Foundation website).

Potential new treatments for CF are rapidly being tested (Korst et al, 1995; Knowles et al, 1995; Davidson and Porteous, 1998; CF Foundation website). These include new anti-microbial agents (such as the promising antibiotic compound IB367), mucolytics, antiproteases, recombinant human secretory leukoprotease inhibitors, and aerosolised α-1 antitrypsin. Delivery of aerosolised amiloride, uridine triphosphate (UTP), or the compound INS37217 to the lung may restore ion transport, and thus improve mucociliary clearance. We are moving increasingly towards treating the causes, rather than the symptoms, of CF lung disease.

Despite these advances, a complete cure for CF remains beyond our grasp. CF is the most common lethal recessive inherited disease, affecting approximately one in 2,500 individuals in European populations. In addition to the personal tragedy, the medical costs of CF average $27,500 annually per patient, and a heart-lung transplant costs in the range of $150,000-200,000 (Korst et al, 1995). One possibility for a cure for CF is gene therapy: by introducing a normal copy of the CFTR gene into epithelial cells of the lung, it might be possible to correct the pulmonary defect.

1.1.4 The CFTR gene

In the quest for new therapies for CF, such as gene therapy, it will be necessary to gain a more complete knowledge of the properties of the CFTR protein and its function within the cell. The CFTR gene lies on the long arm of human chromosome 7 at q31.3, and spans approximately 188.7 kb, encompassing 27 exons, separated by large introns (Fig. 1.1). This gene produces a messenger RNA of 6.5 Kb (Riordan et al, 1989). The predicted open reading frame encodes a 1480-residue glycoprotein of 167 Kda. This protein shows structural homology to the membrane-associated transport ATPase family of proteins (Riordan et al, 1989; Higgins et al, 1990).

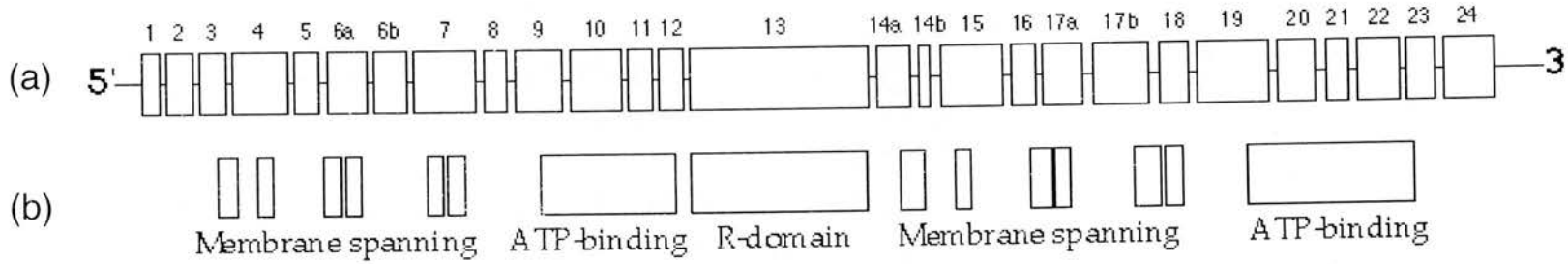


Figure 1.1. Diagram of the CFTR gene. This diagram shows the spatial arrangement of **(a)** Exons 1-24 of the CFTR gene, and **(b)** the corresponding functional domains of the protein (see Fig. 1.2). This diagram was adapted from the CFTR mutation database website. Introns are not to scale.

The CFTR protein is composed of discrete domains:

- 12 hydrophobic membrane spanning domains (TM1 – TM12)
- Two ATP-binding domains (referred to as nucleotide binding domains: NBD1 and NBD2)
- A large 'R domain' which is thought to have a regulatory function (Rich et al, 1997)

Figure 1.2 shows the proposed structure of the CFTR protein (Riordan et al, 1989). The protein consists of two roughly identical halves, each comprised of a set of six membrane spanning regions, followed by an NBD. The large R domain lies in the center, connecting the two halves of the protein (Reviewed in Tsui, 1995; Frizzell, 1995; Davidson and Porteous, 1998). CFTR is an ABC transporter: it is now generally accepted that CFTR is itself the Cl- channel that is lacking in CF. There is some recent evidence that CFTR may form a dimer, although this is debated (Li et al, 2001; Raghuram et al, 2001; Zerhusen et al, 2001).

CFTR Cl- channel activity is tightly regulated. For Cl- transport to occur, cAMP-dependent protein kinases must first phosphorylate sites in the R domain (Rich et al, 1990). ATP hydrolysis of NBD1 produces a state amenable to Cl- transport, while ATP hydrolysis of NBD2 essentially closes the channel. These ATP hydrolysis events do not strictly open and close the channel, but rather increase the open probability, as CFTR rapidly fluctuates between an open and closed state (reviewed in Frizzell, 1995). The R domain may act as a 'ball and chain' to physically block the CFTR channel and prevent Cl- transport during the closed conformation (Winter and Welsh, 1997). Forskolin and other cAMP agonists stimulate secretion of Cl- by the CFTR channel (Bargon et al, 1998), but treatment with phorbol myristate acetate (PMA) prevents this (Trapnell et al, 1991a).

CFTR homologues have been sequenced in mammals such as mouse (Ellsworth et al, 2000) and in organisms as diverse as Fugu (Davidson, H et al, 2000). CFTR is a highly conserved protein: in comparison with the human CFTR gene, sheep CFTR shows 90% identity at the nucleic acid level and 95% similarity at the protein level; mouse CFTR shows 80% identity and 88% similarity; and Fugu CFTR shows 58%

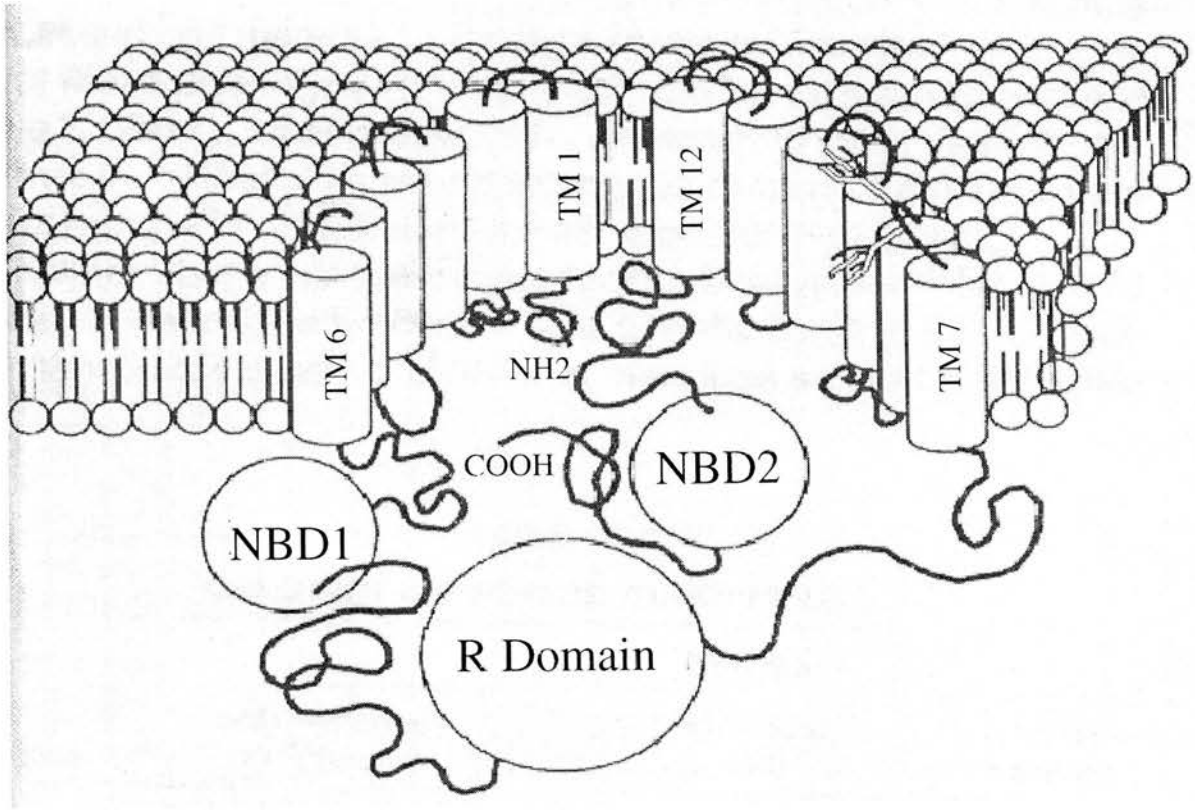


Figure 1.2. Diagram of the CFTR protein. (Modified from Frizzel, 1995) The functional domains of the protein are as follows: **(NBD1), (NBD2)** nucleotide (ATP)-binding domains, **(R Domain)** Regulatory domain **(TM1 to TM12)** transmembrane domains (cylindrical regions), **(NH2)** N-terminus, **(COOH)** C-terminus.

identity and 75% similarity (Tebbutt et al, 1995; Ellsworth et al, 2000; Davidson, H et al, 2000).

In addition to its function as a Cl- channel, CFTR affects the expression of several other genes, including ENaC, a sodium channel (Ismailov et al, 1996), and iNOS (Kelley and Drumm, 1998).

1.1.5 CF mutations

Much work has been done to characterise the mutations responsible for CF (CF mutations database website; Pilewsky and Frizzell, 1999; Mateu et al, 2002; Tsui, 1995). 996 disease-causing mutations have been identified in the CFTR gene. Disease-causing mutations are most prevalent among highly conserved amino acids, particularly in the nucleotide binding domains.

The most frequent mutation, ΔF508, accounts for approximately 66% of the alleles in CF patients. This 3-bp deletion leads to the loss of a phenylalanine residue at position 508 in the protein. Other common mutations (frequency > 1%) include G542X, N1303K, G551D and W1282X.

CF mutations can be separated into five classes, describing their molecular consequences on the CFTR protein:

Class I: No synthesis. In this type of mutation, no CFTR protein is produced. One example is the nonsense mutation G542X (which results in a truncated mRNA).

Class II: Block in processing. In this type of mutation, CFTR protein is not processed correctly to its mature form. This may result in the protein being degraded by proteases, rather than being transported to the apical membrane. ΔF508 is an example of a class II mutation. Experiments have shown that a reduced temperature may allow correct processing of the ΔF508 mutant CFTR, resulting in the formation of a functional channel at the apical membrane (Lukacs et al, 1993). This discovery may have therapeutic potential for class II mutations.

Class III: Block in regulation. With this class of mutation, CFTR shows correct localisation, but defective regulation of Cl- secretion. For example, the missense G551D mutant protein cannot be activated for channel function.

Class IV: Altered conductance. In mutations such as missense R117H, the defective CFTR channel still transports some Cl-, but at a reduced rate. The ion specificity of the channel may also be altered.

Class V: Reduced synthesis. Decreased abundance of protein is a feature of class V mutations. In the alternative splicing mutant 3848 + 10kb C→ T, only a portion of the CFTR mRNA is spliced correctly, resulting in a decreased channel number.

A strong correlation has been demonstrated between the class of mutation and the pancreatic phenotype. The pancreas is normal with 'pancreatic sufficient' (PS) mutations such as R177H. In contrast, mutations such as G551D and ΔF508 affect the pancreas function; these are called 'pancreatic insufficient' (PI) mutations. Other phenotypic effects are correlated with specific mutations, for example, blockage of the meconium ileus occurs in 10% of patients with a ΔF508 mutation, but is less frequent with G551D. However, here is no clear-cut correlation between the mutation and the lung phenotype.

Several other diseases have been associated with mutations in the CFTR gene (CF mutations database; Tsui, 1995). For example, congenital absence of the *vas deferens* is commonly associated with a CFTR mutation. There has been one report of CFTR sequence variations causing an elevation in sweat chloride, in the absence of CF (Mickle et al, 1998). 208 sequence variations have been identified in the CFTR gene, which do not appear to cause disease. Some of these cause amino acid changes which are apparently benign.

Independently segregating modifier genes and environmental factors are likely to have an effect on CF disease severity. Mouse model of CF have proved particularly useful in the dissection of candidate modifier genes (see Section 1.1.11).

A recent report on the phylogeny of CFTR mutations is noteworthy (Mateu et al, 2002). CFTR mutations are most prevalent in Europe, and the allele frequencies display a gradient from the Northeast to the Southwest. The three most common CF mutations, ΔF508, G542X and N1303K, share a common haplotype in the region surrounding the CFTR gene. This suggests that these mutations arose in single population, where this haplotype was common. This haplotype is not found at a high frequency in any of the modern populations that were analysed, which suggests that the expansion of these alleles predates the divergence of modern-day populations. Similarly, the alleles G551D and W1282S share a common haplotype. Populations with a moderate prevalence of this haplotype are widely distributed in multiple continents; this is compatible with a European origin for these mutations.

Theories of heterozygote advantage (for example, against cholera - McDonald et al, 1995) have been proposed to explain the high prevalence of CFTR mutations in the population. Knowing the origin of CF-causing mutations would undoubtedly give us a more rounded picture of the history of the disease.

1.1.6 The mechanism by which CFTR dysfunction produces the lung phenotype

The precise mechanism by which mutations in the CFTR gene cause the severe lung phenotype seen in CF is still unclear. However, several attractive hypotheses have been advanced (Porteous and Davidson, 1997).

The secretory model of CF suggests that the disrupted Cl- channel function of CFTR is directly responsible for the phenotype. Cl- transport is a key component of the salt transport network of the lung; disruption of this function may lead to changes in the airway surface liquid (ASL), with drastic effects on the lung (Guggino, 1999; Pilewsky and Frizzell, 1999). Two contrary hypotheses have been proposed: the hydration hypothesis (Boucher, 1994), suggesting a reduced ASL volume (which may compromise mucociliary clearance), and the salt hypothesis (Smith, JJ et al, 1996), suggesting a hypertonic ASL (which may compromise salt-sensitive antibiotic activity, such as that of defensins).

A decrease in mucous sulfation, along with an increase in sialylation is observed in purified CF mucous, in comparison to non-CF mucous. Mucous sulfation may contribute to the predisposition to bacterial colonization in the CF lung (Zhang et al, 1998b). Alterations in electrolyte transport may affect sulfation; in addition, CFTR may have a direct function as a transporter of sulfate donor molecules (Pasyk and Foskett, 1997).

Other models propose that CF is caused by altered binding of pathogens to the CFTR channel. The internalisation hypothesis suggests that in normal cells, CFTR may act as a ligand for the internalisation of pathogens (such as *P. aeruginosa*), and subsequent cell shedding will rid the lung of these pathogens; thus in CF, effective pathogen binding and cell clearance cannot occur (Pier et al, 1997 and 2000). A contradictory theory suggests that *P. aeruginosa* adheres to CF lung cells at higher levels than to normal cells (due to an upregulation of the asiagloGM1 cell surface protein in CF), and this results in infection by these bacteria (Davies et al, 1997).

Finally, a model of cytokine dysregulation has been proposed to explain the lung defect (Cantin, 1995; Richman-Eisenstat, 1996). Rather than being a secondary response, inflammation may in fact precede infection in CF. Abnormal ion transport or acidification of intracellular organelles might interfere with the expression of anti-inflammatory cytokines. Alternatively, the production of thickened mucus, with altered composition, by the CF submucosal glands might stimulate the release of pro-inflammatory markers such as interleukin-8, to accomplish an inflammatory response in the absence of a bacterial stimulus. Whatever the mechanism, CF lung disease appears to be complex, involving a vicious cycle of infection, inflammation and airway damage.

1.1.7 The CFTR promoter

Much work has been done to elucidate the control elements that drive CFTR expression. The CFTR proximal 5' region seems to have many characteristics of a housekeeping promoter (Yoshimura et al, 1991b) (Fig. 1.3); there is no TATA box, sequences with homology to the CCAAT box have been found at –239 and –123 bp (all positions given are relative to the ATG start codon) and the region from –500 bp

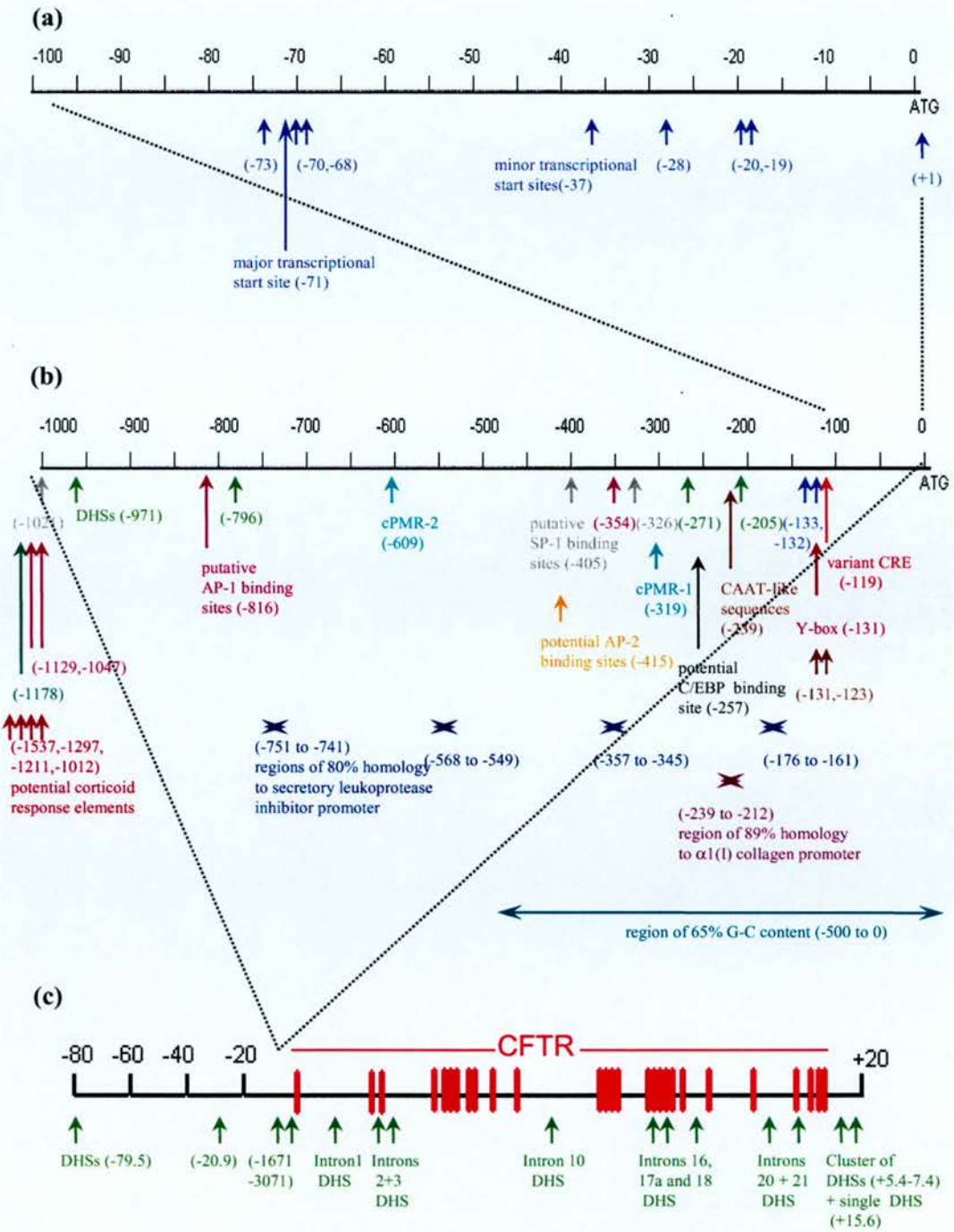


Figure 1.3. CFTR regulatory elements. This is a diagram of the regulatory elements which have been identified in the 5'UTR and intron 1 of CFTR. **(a)** The first scale describes the region from -100 bp to the ATG start codon, and shows various transcriptional start sites.**(b)** The second scale describes the region from -1000 bp to the ATG codon, and shows more transcriptional start sites, putative binding sites, potential response elements, regions of homology, a region of high G-C content, a Y box, a variant CRE, and DHSs. **(c)** The third scale describes the region from -80 kb upstream of CFTR to +20 kb downstream of CFTR, and describes DHSs in the distal 5' region, within CFTR introns and in the 3' region.

to ATG has a high GC content. There are multiple minor transcriptional start sites (at –133, -132, -73, -70, -68, -119, -37, -28, -20, -19 and +1 bp) in addition to the major transcriptional start site at –71 bp. There are also three putative binding sites for the transcription factor SP-1 at –1021, -405 and –326. These are all features of a housekeeping promoter (Koh et al, 1993).

However, investigation of the CFTR expression profile reveals that it is regulated in a tissue-specific manner (see Section 1.1.10). A search for tissue-specific control elements in the 5' region flanking the CFTR gene has identified several interesting features (Fig. 1.3). Chou and colleagues (Chou et al, 1991) claimed to have identified at least three regions regulating CFTR expression by their effect on expression of the CAT reporter gene. Sequences between –27 and –296 bp were capable of driving reporter gene expression. However, an element just upstream of –297 bp repressed this expression. The endogenous CFTR expression levels suggested the presence of another positive regulator upstream of –756, which accomplishes de-repression (however, Chou and colleagues were unable to identify such an element). Two regions of non-random pyrimidine-purine asymmetry (where one strand contains more pyrimidine bases and the other more purine bases – commonly called PMRs) have been found at –319 and –609 bp from the CFTR gene (Hollingsworth et al, 1994); these are perfect mirror repeat elements about a bisecting plane. Similar PMR elements have been implicated in non-B DNA conformations; in addition, an S1 nuclease digestion site (associated with non-B DNA conformations) maps near or within one of these PMR elements in the CFTR gene. This suggests that these PMR elements may play a role in regulating CFTR expression, by altering chromatin conformation, perhaps in a tissue-specific manner.

Matthew and colleagues (Matthews and McKnight, 1996) have confirmed that there is a functional variant cAMP response element (CRE) binding site at –119, along with a Y-box at –131, consistent with the cAMP-dependent regulation of CFTR expression (see Section 1.1.9). Putative AP-2 binding sites exist at –1178 and –415. Finally, a potential C/EBP binding site was found at –257 relative to the ATG start codon. Despite the plethora of putative control elements in the 5' region, coupling of

a 2 kb region 5' of the CFTR gene to the LacZ reporter gene did not fully recapitulate the endogenous CFTR expression profile (Imler et al, 1996).

The mapping of DNase I hypersensitive sites (DHSs) is a powerful technique to dissect transcriptional regulators. DNase hypersensitivity in the chromatin is often (but not always) indicative of a regulatory element. Ann Harris' group has performed extensive mapping of DHS in the CFTR locus. DHS sites have been found in the region 5' of the gene, 3' of the gene, and within CFTR introns.

Both tissue-specific and ubiquitous DHS sites have been mapped in the proximal region 5' of the CFTR gene. DHS sites at –205 and –796 bp are present in multiple cell types, while the sites at –3071, -1671, -971 and –271 are present in HT-29 intestinal epithelial cells, but not in HFL-1 fibroblast cells (Yoshimura et al, 1991a). The two sites have been found in the more distal 5' region (at –79.5 and –20.9)(Smith et al, 1995): these sites did not appear to be tissue-specific (Nuthall et al, 1999b). Deletion of the –20.9 DHS and all upstream DNA in a YAC construct appeared to reduce transgene expression by ~60% in permanent cell lines. In addition, two PMR elements colocalise with the –20.9 DHS (Nuthall et al, 1999b). It should be noted that there is a predicted gene 52.5 kb upstream of CFTR (Ensembl database, Ensembl gene ID ENSG00000154438), the DHS at –79.5 will be within that gene.

A DHS cluster was identified 3' of the gene at 4574 + 5.4-7.4 kb and a single DHS at 4574 + 15.6, these showed some evidence of tissue specificity (Nuthall et al, 1999a). Finally, several DHS were identified in CFTR introns. A DHS identified in intron 1 (185 + 10 kb) almost certainly has a role in regulating expression (Smith, AN et al, 1996). This element augmented reporter gene expression in the Caco-2 and CHO-K1 cell lines, which express CFTR (Smith, AN et al, 1996; Mogayzel and Ashlock, 2000). CFTR YAC transgenics provide further evidence that the intron 1 DHS is involved in regulation (see Section 1.1.11).

DHSs have also been identified in intron 2 (296 + 4.4 kb), intron 3 (405 + 0.7 kb), intron 10 (1716 + 23 kb), intron 16 (3120 + 3 kb), intron 17a (3271 + 0.7 kb), intron

18 (3600 + 7 kb), intron 20 (4005 + 4 kb) and intron 21 (4095 + 7.2 kb) (Smith et al, 2000). The profile of DNase hypersensitivity at these intronic sites varied between different cell lines and human tissue samples, suggesting that they may be involved in tissue-specific regulation. In reporter constructs including the CFTR minimal promoter region and the luciferase reporter gene, inclusion of the intron 20 DHS increased expression by 4.4-fold, and inclusion of the intron 21 DHS increased expression by 1.5-fold (Phylactides et al, 2002).

While it is unlikely that every DHS reported has a role in regulation, these studies form a good starting point for identifying regulatory elements. The presence of these DHS elements suggests that CFTR regulatory signals are complex and are not restricted to the minimal promoter. Elements in the distal 5' region, 3' region, and within CFTR introns may all be involved in determining the CFTR expression profile.

Mutations in the CFTR promoter are rare (Verlingue et al, 1998). However, a few have been identified (CF mutations database website): C → T at -887 bp and T → G at –812 bp (Bienvenu et al, 1995), ΔAGG at –542 bp (Grade et al, 1994), C→ T at –434 bp. Other possible mutations include T → A at –173 bp, G →T at –165 bp and G → A at –104 bp. The mutation T → G at –812 bp is located in a potential AP-1 binding site, and the mutated sequence exhibits an abnormal binding pattern in a gel shift assay, in comparison to wild-type (wt) sequence (Bienvenu et al, 1995). Additionally, a T → G polymorphism (which does not appear to be involved in CF) has been identified at –966 bp.

A 310 kb YAC containing the CFTR locus demonstrated copy-number dependent expression, when integrated into the Caco-2 cell line (Vassaux et al, 1997). This suggests that insulator, or boundary elements may exist in the region flanking the CFTR gene, which are capable of insulating CFTR expression from adjacent regulatory elements.

1.1.8 The βGlobin promoter

In studying the CFTR promoter, it is useful to consider what is known of transcriptional regulation at other loci. The βglobin locus has been well

characterised, and displays an elegant example of transcriptional regulation (Wood, 1996; Chung et al, 1997; Ho and Thein, 2000).

Hemoglobin consists of two αglobin-like chains, and two βglobin-like chains. Alternative βglobin-like chains are employed at different stages of development; these alter the oxygen-binding affinity of hemoglobin. The genes encoding the βglobin-like chains are clustered at a single locus on the long arm of chromosome 11, and are arranged in the order in which they are expressed: ε (foetal) Gγ and Aγ (embryonic), δ and β (adult) (Fig. 1.4a). Expression of these genes is tissue-specific, being restricted to erythroid cells.

A locus control region (LCR), consisting of five DNase I hypersensitive sites (each associated with regulatory elements), lies upstream of the gene cluster, from –5 to –25 kb relative to the ε ATG start codon. The LCR affects expression in two ways. Firstly, it produces an open chromatin (euchromatin) conformation at this locus, conferring position-independent expression (e.g. unaffected by adjacent heterochromatic regions which would silence expression). Secondly, the LCR acts as a developmental stage-specific enhancer, regulating the switch in expression between the different βglobin-like genes. The favored theory for the mechanism of this enhancement is that a 'holocomplex' forms in the DNA, allowing the LCR to associate with the proximal promoter of one of the βglobin-like genes, with the intervening DNA looping out (Fig. 1.4b). This interaction appears to be governed by the binding of developmental stage-specific, erythroid-specific transcription factors to conserved elements within the LCR (CACCC, [T/A]GATA[A/G], and TGA[C/G]TCA) and the βglobin-like genes' proximal promoters (e.g. TATA, CCAAT and CACCC in the β proximal promoter). For example, erythroid krüppel-like factor (EKLF) binds to the CACCC box. Evidence suggests that binding of the LCR to the βglobin-like promoters is competitive, such that only one gene may be transcribed at a time.

Thus, regulation at the βglobin locus is tightly regulated through development in a tissue-specific fashion. It will be interesting to discover whether similar mechanisms of regulation are employed at the CFTR locus.

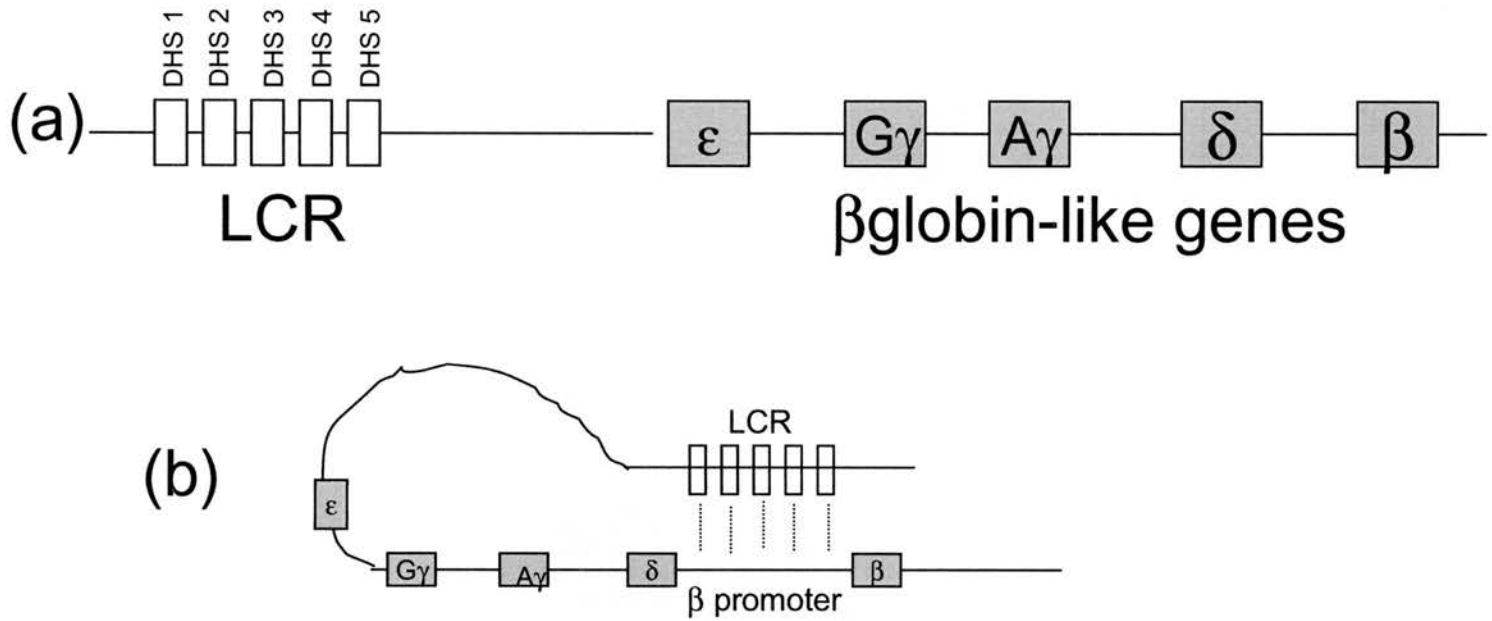


Figure 1.4. The βglobin locus. (a) The βglobin locus consists of five βglobin-like genes: ε (foetal), Gγ and Aγ (embryonic), δ and β (adult). The locus control region (LCR) lies 5 to 25 kb upstream of the genes, and contains five DNase I hypersensitive sites (DHS1-5). **(b) The holocomplex.** The locus control region appears to associate with the proximal promoter of one of the βglobin-like genes to induce transcription, causing the intervening DNA to 'loop-out'. This figure was adapted from Wood, 1996.

1.1.9 Modulation of CFTR expression by pharmacological agents

CFTR expression is affected by several pharmacological agents. Forskolin appears to increase CFTR mRNA levels in HT-29 cells, but conversely, lowers CFTR mRNA levels in T84. This effect appears to be a result of the cAMP-agonist effect of forskolin: membrane-permeable cAMP analogues (CPT-cAMP, DB-cAMP and 8-bromo-cAMP) and cholera toxin (a cAMP agonist) have a similar effect to forskolin, while ATP, non-membrane permeable cAMP, and dideoxy-forskolin (a forskolin analogue which does not affect adenosine cyclase) do not affect expression (Bruer et al, 1992; Bargon et al, 1998). Most interestingly, cAMP treatment has been shown to correlate with an increase in DNase I hypersensitivity at the DHS overlapping the Y box (Pittman et al, 1995). PMA reduces CFTR expression in both the HT-29 and T84 cell lines (Bruer et al, 1992; Trapnell et al, 1991a). Inhibitors of protein kinase A (PKA) suppress even basal levels of CFTR mRNA expression, suggesting that PKA is required for CFTR transcription (McDonald et al, 1995). Finally, extracellular hyposmolarity, or the inclusion of urea or mannitol in growth medium reduce CFTR mRNA expression (Baudouin-Legros et al, 2000).

It is interesting that cAMP agonists and PMA also affect CFTR at the protein level, altering the channel's Cl- gating (see Section 1.1.4). These agents appear to regulate Cl- transport by multiple mechanisms.

1.1.10 CFTR expression profile

CFTR appears to contain a plethora of regulatory elements, and indeed the gene's expression follows a complex pattern, both temporally and spatially. The CFTR expression profile has been characterised by reverse transcriptase polymerase chain reaction (RT-PCR), *in situ* hybridisation, and immunohistochemistry. The latter has been complicated by the difficulty to raise specific antibodies to the CFTR protein.

1.1.10.1 The gut

CFTR protein is abundant in epithelial cells of human intestinal crypts. Expression appears to be restricted to the apical portion of the cells, and a proportion of the protein is found to be associated with the plasma membrane. In the jejunum, expression appears to be highest in the mucosal epithelial cells, although lower levels

are seen in the epithelial cells lining the crypts of the large intestine (Crawford et al, 1991). In rat intestine, CFTR shows decreasing gradients of expression along both the crypt-villus and proximal-distal axes (Trezise and Buchwald, 1991).

1.1.10.2 The pancreas

CFTR protein is expressed in epithelial cells of the pancreatic secretory ducts. This expression is restricted to the smaller of the secretory ducts, and to the luminal portion of these cells (Crawford et al, 1991).

1.1.10.3 The respiratory system

It has been difficult to characterise the CFTR expression profile in the lung: such studies have been frustrated by high levels of background staining when performing immunohistochemistry on this tissue. RT-PCR has demonstrated that CFTR is expressed at ~1-2 mRNA copies per cell in human nasal, tracheal, and bronchial epithelium, in contrast to pharyngeal epithelium, where levels are 100-fold lower (Trapnell et al, 1991b).

Submucosal glands are the main site of CFTR expression in the lung (Engelhardt et al, 1992). Fig. 1.5 shows a cartoon of the lung, including a diagram of the submucosal gland and the pseudostratified airway epithelium. Staining was seen in both the serous components of the secretory tubules (where almost every cell showed high levels of CFTR protein localizing to the apical surface) and in subpopulations of both columnar and basal cells in the ciliated ducts. Expression was not as marked in the surface respiratory epithelium of the proximal airways, and was restricted to a small population of non-columnar cells.

A later study (Engelhardt et al, 1994) detected lower levels of CFTR expression in the more distal airways (including the proximal, terminal, and respiratory bronchioles, and alveoli). In the proximal bronchioles, this expression was localised to the apical surface in a population of non-ciliated cells (these co-localise with the Clara cell marker CC10). Expression was more random in the more distal airways and alveoli.

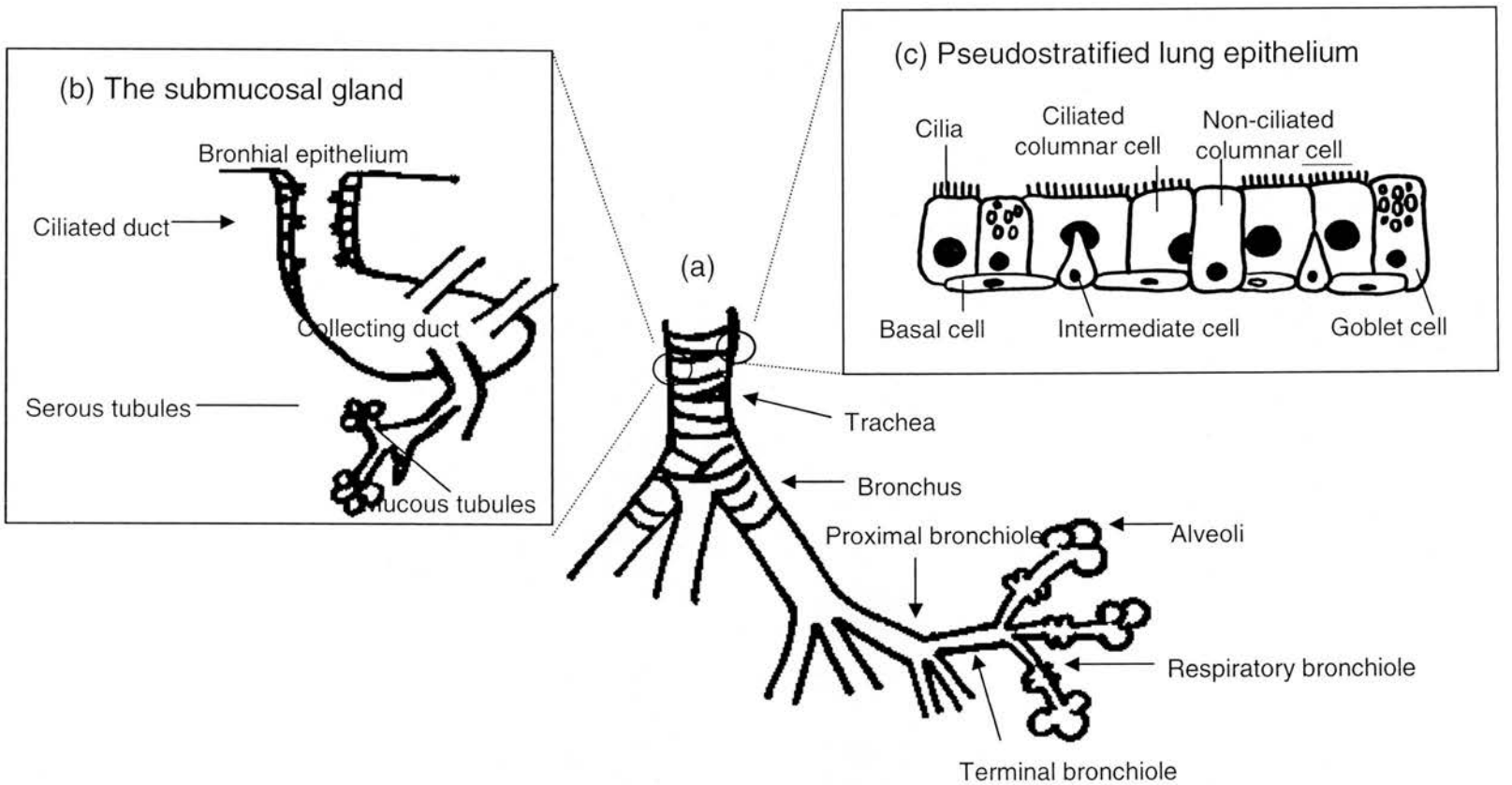


Figure 1.5. A cartoon of the lung. This diagram shows **(a)** the structure of the branching respiratory system, including the trachea, the bronchus, the proximal, terminal and respiratory bronchioles, and the alveoli (as labelled) (adapted from Engelhardt et al, 1994), **(b)** the submucosal gland including the bronchial epithelium, the ciliated and collecting ducts, and the mucous and serous tubules (as labelled) (adapted from Engelhardt et al, 1992) and **(c)** the pseudostratified lung epithelium, including ciliated and non-ciliated columnar cells, intermediate cells, basal cells and goblet cells (adapted from Zhang et al, 1998b).

1.1.10.4 Other organs

The infertility in CF males and reduced fertility in CF females would predict a role for CFTR in the reproductive system. Indeed, CFTR expression has been detected in the seminiferous tubules of males and the uterus epithelial lining in females, in rat tissues (Trezise and Buchwald, 1991).

CFTR protein is abundant in epithelial cells lining the sweat ducts (Crawford, 1991). Kidney tubules show a high level of CFTR protein, despite the fact that the kidney is not affected in CF (Crawford, 1991).

CFTR mRNA is expressed in the heart, and CFTR appears to have a function as a Cl- channel in this tissue (Levesque et al, 1992), although the heart does not appear to be affected in CF. There is an epicardial to endocardial gradient of CFTR expression in the left ventricle of the rabbit heart (Wong et al, 1999), which may represent an alternatively spliced form of CFTR (Horowitz et al, 1993).

1.1.10.5 Non-epithelial cells

Early investigations suggested that CFTR expression was restricted to epithelial cells (Crawford et al, 1991). However, there is now some evidence of expression in other cell types (Yoshimura et al, 1991a; Tousson et al, 1998). There is evidence of CFTR expression in lymphocytes, sertoli cells, heart muscle cells, non-epithelial cells of the submucosal gland, and hypothalamic neurons. CFTR expression has been demonstrated in mast cells, and it seems likely that CFTR may be involved in mast-cell activation and secretion, through its Cl- transport function (Kulka et al, 2002).

CFTR expression appears to be regulated during development. Studies of fetal tissue from pancreas, intestine, and vas deferens of CF patients show developmental abnormalities For example, dilation or destruction of the pancreatic glands and ducts, and abnormalities or absence of the epididymus and vas deferens in males (Zuelzer and Newton, 1969; Thomaisis and Arey, 1963; Valman and France, 1969, Ornoy et al, 1987).

CFTR is expressed in the bronchial epithelium during the first and second trimesters, with a gradient of expression diminishing from the bronchi, to the bronchioles, to the pre-alveolar tubes. This expression appears to be uniform in the first trimester, but becomes restricted to patches in the second trimester, perhaps reflecting cell differentiation (McCray et al, 1992). The fetal pulmonary lung secretes fluid: this process is thought to be driven by Cl- secretion and to be involved in development. Thus, there is potential for CFTR to have a role in lung development. One study has claimed to reverse the phenotype in a CF mouse model by *in utero* gene therapy (Larson et al, 1997). In this study an adenovirus vector was administered to the CF mouse *in utero*. This appeared to drive CFTR expression during development, but expression became switched off after birth. They claimed that this completely rescued the intestinal phenotype of the S489X CF mouse; this would suggest that it is CFTR's role in development that causes the gut phenotype. However, this data is controversial. In addition, the lack of obvious morphology in CF neonate lungs casts doubt on the hypothesis that it is CFTR's role in development that is responsible for the respiratory phenotype in CF patients.

1.1.10.6 ΔF508 CFTR expression in CF

mRNA levels were found to be similar in the respiratory epithelium from normal and ΔF508 CF patients (Trapnell et al, 1991a). However, *in vitro* studies (Lukacs, 1993), and later, *in vivo* studies (Kartner et al, 1992) have shown that the ΔF508 protein fails to localise to the apical membrane. In addition, at the protein level, ΔF508 shows a different distribution to wt CFTR: while similar levels of protein are seen in respiratory and intestinal tracts, ΔF508 protein is undetectable in the sweat gland (Kälin et al, 1999).

1.1.11 Animal models of CF

Animal models are important tools in biological research. The possibility of creating knockouts and transgenics in the mouse has opened new doors of understanding for many diseases.

Several groups have generated mouse models of CF (Table 1.1) (reviewed in Davidson and Dorin, 2001). These can be broadly separated into three classes: (1)

compete knockouts, (2) 'leaky' knockouts (where a low level of functional CFTR is produced, for example, due to alternative splicing of insertion mutants), and (3) models of specific clinical mutations, for example, ΔF508 and G551D (interestingly, all but one of the 60 clinical missense mutations are in sequences conserved between the human and mouse).

Table 1.1. Mouse models of CF.

Mouse Model	Class	Lesion	Reference
$CFTR^{tm1Unc}$	1	Exon 10 replacement	Snouwaert et al, 1995
$CFTR^{tm1Cam}$	1	Exon 10 replacement	Ratcliff et al, 1993
$CFTR^{tm3Bay}$	1	Exon 2 replacement	Hasty et al, 1991
$CFTR^{tm1Hsc}$	1	Exon 1 replacement	Rozmahel et al, 1996
$CFTR^{tm1HGU}$	2	Exon 10 insertional	Dorin et al,1992
$CFTR^{tm1Bay}$	2	Exon 3 insertional duplication	O'Neal et al, 1993
$CFTR^{tm2Cam}$	3	ΔF508 by exon 10 replacement	Colledge et al, 1995
$CFTR^{tm1kth}$	3	ΔF508 by exon 10 replacement	Zeiher et al, 1995
$CFTR^{tm1Eur}$	3	ΔF508 by exon 10 'hit and run'	Doorninck et al, 1995
$CFTR^{tm1G551D}$	3	G551D by exon 11 replacement	Delaney et al, 1996
$CFTR^{tm2HGU}$	3	G480C by exon 10 'hit and run'	Dickinson et al, 1998

(This table has been adapted from Davidson and Dorin, 2001)

All of these mice show a defect in Cl- transport in epithelial cells, characteristic of CF (Clarke et al, 2002). Varying degrees of intestinal phenotype are displayed, according to the genetic lesion, and the mouse strain. However, none of the mouse models shows convincing evidence of lung disease under normal housing conditions. The sterile living environment, short life span, and possibly the presence of an amiloride-insensitive, electrogenic Na+-glucose transport system in the mouse airway (which is not present in humans) (Joris and Quinton, 1989) may account for the failure of the mouse to develop a spontaneous morbid lung phenotype. At least two groups have been successful in demonstrating a differential immune response and bacterial clearance between CF mutant mice and their wild-type littermates, following a forced infection with pathogens commonly found in CF patients (Heeckeren et al, 1997; Davidson et al, 1995).

YAC transgenics have also been informative. A transgenic mouse, carrying a YAC containing the intact human CFTR locus was made (Manson et al, 1997). This YAC

transgenic shows an expression pattern similar to endogenous CFTR (where there is a difference in expression patterns between mouse and human CFTR, the transgene appears to follow the mouse pattern). This YAC transgene is capable of rescuing the intestinal phenotype of the $CFTR^{tm1Cam}$ knockout: YAC transgenic x $CFTR^{tm1Cam}$ mice appear phenotypically normal and breed well, in contrast to the null mice, 50% of which die perinatally. To test the effects of the intron 1 element upon expression, A YAC was generated where the intron 1 DHS element had been deleted: this YAC was used to create another transgenic mouse. Deletion of the intron 1 DHS was shown to reduce expression in the intestine by 60%, in contrast to the YAC transgenic with intact intron 1, while expression in the lung was similar between the two constructs (Rowntree et al, 2001). This provides further evidence that the intron 1 DHS element accomplishes upregulation of CFTR expression in the gut.

Although the inbred strains used to make transgenics may be a poor model of the heterogeneous CF population, these mouse strains have proved useful in the dissection of modifier genes. A major modifier locus has been mapped in the mouse near the centromere of chromosome 7, which produces a 'prolonged survival' phenotype in $CFTR^{tm1Hsc}$ null mice (Rozmahel et al, 1996), by strain interbreeding.

1.1.12 Larger model organisms

Larger mammals may provide better models for CF studies. The sheep, in particular, has a number of advantages (Harris, 1997; Traber and Traber, 1989):

- Sheep are of similar size to humans.
- Sheep have a life span close to that of humans, this may be necessary for the development of the complex respiratory phenotype seen in human CF. The longer life span will also allow long-term studies to be performed.
- Sheep and humans have a similar cardiopulmonary system (Harris, 1997).
- The cloning of Dolly the sheep (Wilmut et al, 1997) and genetic engineering of Polly, a transgenic sheep, (Schnieke et al, 1997) have opened up the possibility of creating a sheep model of CF by targeting the CFTR gene for knockout or mutation.

- Because of their similarity to humans, the sheep may develop a lung phenotype that closely mimics CF, where the mouse has not.

However, the sheep has a number of disadvantages as a model organism, as well:

- Sheep are more expensive to keep, and larger volumes of reagents will be required to treat the larger sheep tissues.

- Inbred sheep strains do not exist and one would expect modifier genes to be present in the outbred population (in contrast to the genetically homogenous mouse strains). While this may provide a more realistic model of the human condition, these modifier genes may complicate the phenotype and increase the variability of experimental results.

- Conversely, the longer life span of the sheep means that scientists will have to wait longer for animals to reach maturity, in order to set up breeding programs and utilise these adult animals for experiments. For example, humans with CF are usually diagnosed at four years of age, thus once the first CF sheep is produced it may be four years before we can establish whether it displays a respiratory phenotype.

- The non-sterile living conditions in which sheep are currently housed may lead to complications.

- There are additional ethical issues in sheep experiments, and it is more difficult to obtain a license for such experiments.

It seems likely that both the traditional mouse models and larger mammal models will play a role in future studies of Cystic Fibrosis.

1.2 Gene Therapy for Cystic Fibrosis

1.2.1 Rationale behind gene therapy for CF

One possibility for the treatment of cystic fibrosis is gene therapy: by introducing a normal copy of the CFTR gene into epithelial cells of the lung, it may be possible to direct expression of functional CFTR protein, and thus correct the respiratory defect

(Beardsley, 1990). Several features of CF make this disease particularly amenable to gene therapy (Douglas and Curiel, 1998; Alton et al, 1993). CF represents a significant cause of mortality, for which there is no known cure, justifying the development of additional therapies (Gaiso, 1998). The disease is monogenetic and the genetic lesion is well defined, providing a clear target: augmenting cells with a normal copy of the CFTR gene should be sufficient for a cure.

The human lung has a surface of 1-2 square meters, with successively branching 'fractal-like' structure (Fig. 1.5). The stem cells of the respiratory epithelium have not been precisely identified, and it will undoubtedly be difficult to access these cells for removal. Hence, an *ex vivo* approach would probably not be suitable (Rosenfeld et al, 1992; Midleton and Alton, 1998; Flotte et al, 2001). However, the airway is relatively accessible for *in vivo* transfer: aerosol delivery of medicines successfully treats respiratory diseases such as asthma. Nebulisation of the lung is an alternative delivery method that may produce more efficient gene transfer to the respiratory epithelium (Flotte et al, 2001).

The precise mechanisms by which CFTR mutations cause the debilitating lung phenotype in CF remain unclear (see Section 1.1.6). It seems unlikely that the alveoli will be the therapeutic target (Korst et al, 1995). The submucosal glands are the predominant site of expression in the lung. However, high levels of expression do not guarantee that these are the affected cell types: both the kidney and the heart express high levels of CFTR, but do not display a phenotype (see Section 1.1.9). If submucosal glands turn out to be the required target for CF gene therapy, it is unlikely that airway administration will be successful: these cells might be more successfully transfected by intravenous administration coupled with a method to target this cell type (Middleton et al, 1998). However, the surface respiratory epithelium of the proximal airways is the current focus as a target for gene therapy studies, and these cells are relatively accessible for airway administration.

1.2.2 Gene Therapy vectors

The vectors used in gene therapy fall into two classes: viral and non-viral. Both approaches have been considered to treat CF (reviewed in Davies et al, 1998).

1.2.2.1 Viral vectors

Viral vectors exploit the natural ability of viruses to infect cells and drive expression of their genetic material. For safety reasons, most viral vectors are made replication-deficient, by removing some, or all of the viral genes required for replication. The first gene therapy vectors were based on retrovirus. However, these vectors only infect dividing cells, and the respiratory epithelium displays a low level of proliferation. In addition, retroviral vectors are unstable in the presence of serum and it is difficult to produce these vectors in high titres. Thus, retroviral vectors may be a poor choice for CF gene therapy.

More promising viral vectors are based on adenoviruses, which can infect cells that are not proliferating, and have a natural tropism for the respiratory epithelium (Rosenfeld et al, 1992; Crystal et al, 1994; Imler et al, 1996; Davies et al, 1998). Adenovirus is thought to enter the cell by receptor-mediated endocytosis. The virion then ruptures the endosome, gaining entry into the cytoplasm. Adenovirus infection is not associated with malignancy, as the viral DNA does not integrate into the host genome, eliminating the risk of insertional mutagenesis. Natural adenovirus produces an immune response, which could be particularly problematic for CF patients. In addition, infection is usually short lived, often resulting in attenuation, thus preventing administration of the same vector twice. Finally, adenovirus does not infect columnar, and more differentiated cells of the airway: these may turn out to be the therapeutic targets for CF. In first-generation adenoviral vectors, the early immediate genes E1a, E1b and E3 were deleted, to prevent viral replication, these vectors produced an immune response (see Section 1.2.5). In second-generation adenoviral vectors, the E2a and E4 regions have been deleted, this appears to reduce the immune response and may produce longer-lived expression. In third generation adenoviral vectors, no viral genes remain: this should combat the issue of cell-related immune responses. However, the problem of humoral immune response to the viral caspid remains (Flotte et al, 2001).

Adeno-associated viruses (AAVs) might also be used as gene therapy vectors (Douglas and Curiel, 1998). These share many of the advantages of adenoviruses, for example, the tropism for the respiratory tract and the ability to infect non-dividing

cells. Unlike adenovirus, AAVs may transduce multiple cell types. This type of vector allows integration of DNA into the host genome, which may allow for longer-lived expression, but conversely produces the possibility of insertional mutagenesis. AAV virus requires adenovirus or herpes virus to replicate, and thus AAV vectors should be naturally replication deficient (Flotte et al, 2001). Like retrovirus, AAV can only be produced in low titers, and AAV transgene accommodation is limited (4.5 kb maximum). AAV may not induce the immune response seen with adenovirus.

1.2.2.2 Non-Viral Vectors for gene therapy

Non-viral vectors, such as cationic liposomes, provide an alternative vehicle with increased safety and reduced immunogenicity in contrast to viral vectors (Douglas and Curiel, 1998; Goddard, 1997). These vectors accomplish DNA transfer by membrane fusion, endocytosis, or other mechanisms. Non-viral vectors do not have the advantage of millions of years of evolution to maximise entry into the cell, but conversely, the cell has not had millions of years of evolution to adapt to preventing their entry.

Bacterial plasmids are the primary vehicle used in non-viral gene therapy research. These circular DNA molecules can be engineered to carry the gene of interest and can be produced in bulk. In most circumstances, plasmids would remain episomal following gene transfer. The plasmid will not undergo replication, thus repeat administrations will be required as the therapeutic gene is diluted with successive cell cycles. However, alternative approaches are being considered. For example, several groups are investigating a 'gene correction' approach, where the endogenous CFTR gene is targeted for repair through recombination with a primer, short DNA fragment, or RNA/DNA hybrid 'chimeroplast' (Thorpe et al, 2002). Another alternative is to create a human artificial chromosome (HAC) as a gene therapy vector. This would be maintained in a similar fashion to an endogenous chromosome (Saffery and Choo, 2002).

Initial experiments testing transfection of naked plasmid DNA into epithelial cells have met with limited success (Holmes et al, 1999) Transfection reagents have the potential to increase the efficiency of gene transfer. For example, liposomes (such as

DOTAP), Saint-Mix™, Polyethylenimine, or the LID method may enhance transfection (see Section 4.1 for more on these reagents).

In vivo, expression from plasmid vectors appears to be short-lived, beyond that accounted for by dilution of the transgene through the cell cycle. This may be due to inactivation of the transgene, exclusion of the transgene from the cell, or shedding of transfected cells. This may be a specific response following recognition of certain bacterial DNA sequences in the plasmid. Preliminary experiments have shown a lack of attenuation to cationic liposome-DNA complexes, allowing the possibility of repeat administration to prolong expression.

1.2.3 In vitro and ex vivo Experiments

In vitro experiments as early as 1990 have shown that introducing the normal human CFTR cDNA into cultured cells drives cAMP-mediated Cl- secretion. Drumm and colleagues (Drumm et al, 1990) used amphotrophic retrovirus to transform CFTR cDNA into the CFPAC-1 cell line (a pancreatic adenocarcinoma line from a CF patient, which does not express CFTR). cAMP-dependent Cl- secretion was demonstrated by measuring intracellular I^{125} efflux, and whole-cell patch-clamp technique, in transformed cells; this activity was not present in untransformed controls. Gel blot hybridisation showed the presence of CFTR DNA in these cells, and RNA blot hybridisation demonstrated the presence of CFTR mRNA. Hence, it seems likely that expression of CFTR from the vector produced a functional CFTR channel, responsible for the restoration of normal ion transport properties to the cells. In addition, Rich and colleagues (Rich et al, 1990) showed that vaccinia-viral mediated transfer of normal CFTR cDNA, but not ΔF508 CFTR cDNA, to cultured CF airway cells restored normal electrolyte transport properties, as assayed by the patch-clamp technique.

In order to study the percentage of cells that must express normal CFTR to achieve functional correction, Johnson and colleagues (Johnson et al, 1992) transfected CF airway cells with either CFTR or a reporter gene and mixed the cells in different proportions, before seeding them to form confluent epithelial sheets. They found that epithelial sheets containing only 6-10% corrected cells (expressing CFTR) had Cl-

transport properties similar to sheets containing 100% corrected cells. They speculated that this might be due to movement of Cl- between cells in the epithelial sheet, through gap junctions. This suggests that it might be possible to restore normal lung function by correcting only a proportion of the cells in the lung with gene therapy.

Finally, Zhang and colleagues (Zhang et al, 1998b) compared a recombinant adenoviral vector with liposomal transfection of a plasmid expressing CFTR, to asses the correction of two defects in a CF bronchial xenograft model: Cl- transport and mucous sulfation (see Section 1.1.6). They found that the adenoviral vector efficiently transduced cells, driving a high level of CFTR transgene expression, and correcting the Cl- transport defect to 91% of non-CF levels. However, the mucous sulfation defects were not corrected, apparently because adenovirus does not target goblet cells. In contrast, liposomal transfection of a CFTR-encoding plasmid produced lower expression levels, and only corrected Cl- transport to 7.4% of non-CF levels. However, liposomal transfection efficiently reduced mucous sulfation. These findings reveal that different vectors have different complementation profiles, and that a low level of CFTR expression may suffice to correct some of the CF airway defects.

1.2.4 Animal experiments

The development of CF mouse models has been a watershed in the study of CF, and these models have played a key role as an intermediary between *in vitro* experiments and human trials of gene therapy (see Section 1.1.11). Delivery of a CFTR plasmid/liposome complex to the respiratory epithelium of mice resulted in CFTR mRNA expression in the lung (including large and small airways). Delivery of this complex to the airways of $CFTR^{tm2Cam}$ mice restored the defect in cAMP-stimulated Cl- secretion in the tracheae of these mice (Hyde et al, 1993). In a similar fashion, the DOTAP liposomal reagent was successful in restoring the electrophysiological defect in CF mutant mice (McLachlan et al, 1996). Alton (Alton et al, 1993) showed that delivery of human CFTR cDNA-DC-Chol/DOPE liposomal complex by nebulisation to the airways of the $CFTR^{tm1HGU}$ mouse restored the presence of CFTR mRNA, at a range of levels (averaging 50% of wt levels). This was superior to

instillation of the complex, which only produced modest correction. A restoration of Cl- transport properties was also demonstrated in the nasal and tracheal epithelium, by measuring potential difference (PD)[1].

Dorin and colleagues (Dorin et al, 1996) crossed mice with the $CFTR^{tm1UNC}$ allele and the $CFTR^{tm1HGU}$ allele, to produce mice that expressed different levels of CFTR. The $CFTR^{tm1UNC}$/ $CFTR^{tm1HGU}$ compound heterozygote expressed 5% of normal levels, while $CFTR^{tm1UNC}$/wt, expressed 50%, and $CFTR^{tm1HGU}$/wt expressed 55%. This experiment demonstrated a non-linear relationship between phenotype and gene activity: they found that 5% of the normal levels of CFTR mRNA were sufficient to restore 50% of the normal Cl- transport, and essentially rescue the intestinal disease phenotype. This further justifies human gene therapy trials, as it lends strength to the hypothesis that only a fraction of gene activity must be restored to the lung to achieve normal lung function. In sum, these experiments have provided a strong premise for gene therapy trials in CF patients.

1.2.5 Clinical Trials

Human gene therapy trials for CF include three successive phases: phase I trials test the safety of the therapy, phase II trials test efficacy in a small groups and phase III trials show efficacy in a larger group, confirming the suitability of the therapy for regular use. CFTR delivery has been tested in phase I trials in the nasal and lung epithelia.

---

[1] PD measures the change in voltage, which is affected by the flow of ions across the cell membrane. cAMP-dependent chloride secretion can be measured by first administering amiloride, which blocks sodium channel activity, causing a reduction in the levels of potential, then administering forskolin, which stimulates cAMP, and thus acts as an agonist for CFTR Cl- secretion (see Section 1.1.4). In CF cells, there is no response in PD following the administration of forskolin, as no chloride flow through CFTR can be induced, whereas normal cells respond with an increase in potential difference.

A trial by Zabner and colleagues (Zabner et al, 1993) demonstrated transfer of a replication deficient adenovirus to the nasal epithelium of CF patients. This tissue was chosen for the trial because it had a morphology and function similar to the lung epithelia and manifested the CF Cl- transport defect. In addition, the nasal epithelium was more accessible than the lung and minimised the risk to the patient if an adverse reaction should occur. Gene delivery resulted in a partial restoration of Cl- transport, as measured by nasal voltage change following the addition of amiloride and then cAMP agonists. In addition, CFTR transcripts were detected by RT-PCR. In this trial, there was no evidence of viral replication or virus-associated adverse effects.

In another study by Crystal and colleagues (Crystal et al, 1994), adenoviral vector expressing the CFTR cDNA sequence was introduced into the nasal and bronchial epithelium of CF patients. Preliminary experiments in animals had demonstrated no change in clinical safety parameters, although there was a moderate inflammatory response. Nasal administration to CF patients produced no adverse response. Adenoviral CFTR DNA and mRNA were detected in nasal brushings, and CFTR protein was detected in some cells by antibody staining. However, following bronchial administration to CF patients, a substantial transient systemic and pulmonary syndrome was observed. This serious side effect has cast doubt on the suitability of viral vectors for CF gene therapy.

Many CF gene therapy investigators have responded to this adenoviral-induced immune response by switching to a non-viral vector. Caplen and colleagues (Caplen et al, 1995) attempted to deliver a liposome/CFTR complex to the nasal epithelium. PCR products (specific to the introduced CFTR copy) were amplified from nasal biopsies. RT-PCR confirmed the presence of a CFTR transcript specific to the vector in samples from several patients. *In vivo* nasal PD demonstrated some restoration of normal ion transport properties, with some measurements reaching the non-CF range. Most importantly, there was no evidence of treatment-related toxicity. Similar studies by Gill and colleagues (Gill et al, 1997), and Porteous and colleagues (Porteous et al, 1997) have demonstrated transgene expression and restored Cl- transport in the nasal epithelium following treatment with liposome/CFTR plasmid

complexes, with no related immune response. Hence, non-viral vectors may provide a better prospect for further gene therapy trials.

However, in a later trial by Alton and colleagues (Alton et al, 1999), administration of a CFTR-cDNA/GL-67 liposomal complex to the nasal or bronchial epithelium of CF patients caused mild, influenza-like symptoms in seven out of eight patients receiving the treatment, and six out of eight patients in both the treatment and placebo groups reported mild airway symptoms. While both effects were mild and resolved spontaneously, this underlines the need to investigate safety parameters and refine gene therapy treatments to non-toxic formulations, even amongst non-viral therapies. The CFTR-cDNA/GL-67 treatment produced a significant correction in the Cl- transport properties (as measured by PD and chloride efflux) and bacterial adherence was reduced.

1.2.6 Hurdles to CF gene therapy

Early studies have identified a number of difficulties directly relating to gene transfer (Davies et al, 1998; Flotte et al, 2001). The cell has evolved multiple barriers to prevent entry of foreign particles. A successful vector for gene therapy must show proficiency in:

- Crossing the mucus at the cell surface
- Crossing complex cell surface structures such as cilia
- Achieving cell entry (e.g. by endocytosis)
- Escaping the endosome, to enter the cytoplasm, if necessary
- Entering the nucleus
- Driving CFTR expression

If gene therapy for cystic fibrosis is to become a reality, several issues will need to be resolved. Most importantly, there is the issue of safety: the vector must not induce an adverse immune or other morbid response in the patient. In addition, if a viral vector

is used, the possibility that this vector may somehow reproduce and infect other tissues in the patient or other individuals, especially the germ-line, must be guarded against. Clinical studies of CF gene therapy are facilitated by the fact that CF males are infertile (see Section 1.1.2), thus germ-line gene transfer is not as problematic. As most current vectors for gene therapy show only transient expression, it will be necessary to alter the vector for more prolonged expression, or design a vector which is suitable for repeat administration. Attenuation to re-infection is a feature of viral vectors, however, Goddard and colleagues have shown that a repeat dose of liposome is as affective as the first in delivering CFTR cDNA to the mouse trachea, and in correcting cAMP-dependent Cl- secretion in this tissue (Goddard et al, 1997).

Although low levels of expression and expression in a small portion of the cells may be sufficient to correct the defect in Cl- secretion, a complete restoration of the ion transport properties (e.g. correction of the sodium hyperabsorption characteristics) may require a more comprehensive restoration of gene activity (Douglas and Curiel, 1998). Thus, the efficiency of transgene delivery and expression may have to be increased by the development of improved vectors and delivery systems. Finally, there is evidence that expression of CFTR at excess levels (Stutts et al, 1993; Wei et al, 1995; Schiavi et al, 1996, Mohammed-Panah et al, 1998), or in inappropriate cell types (Ye et al, 2001) is deleterious. More research is required to elucidate the affected cells that cause CF lung disease, and to target these cells in gene therapy, either at the level of cell entry or at the level of gene expression. A logical solution would be to use sequences from the natural CFTR promoter to drive gene expression, as opposed to the heterologous constitutive promoters (usually derived from a virus), which have been used in most gene therapy trials so far (Boyd et al, 1999b): this is the rationale behind 'genomic context vectors' (GCVs) for CF gene therapy.

In the quest for better gene therapy vectors, it will be necessary to use reporter genes to monitor expression. This report investigates one such reporter gene: Enhanced Green Fluorescent Protein (EGFP).

1.3 EGFP as a reporter gene

Previously, reporter genes such as LacZ (Imler, 1996; Boyd et al, 1999a), luciferase (Smith, AN et al, 1996a; Mogayzel and Ashlock, 2000), and CAT (Chou et al, 1991) have been used to assay expression from the CFTR promoter. New reporter gene constructs, based on the Green Fluorescent Protein (GFP) may prove to be a valuable supplement to this list.

1.3.1 The GFP gene

Green Fluorescent Protein was isolated from the jellyfish *Aquorea victoria* and is a naturally fluorescing protein, which absorbs light in the blue region of the spectrum (~470 nm), and emits light in the green region of the spectrum (with a peak at ~509 nm, and a shoulder at ~540 nm). The spectral properties of GFP are achieved by a hexapeptide repeat, which starts at the 64^{th} amino acid of the GFP protein. The primary sequence of the protein in this region results in cyclisation of serine-dehydrotyrosine-glycine, producing an active chromophore (or fluorophore), which produces the fluorescence (Stauber et al, 1998). The protein forms a cylinder, made up of 11 strands of β-sheet on the outside, and an α-helix on the inside; GFP was the first protein to be identified with this novel arrangement of domains, named a 'β-can' (see Fig. 1.6). The fluorophore is contained on the inside of the cylinder (Yang et al, 1996).

GFP offers several advantages over more traditional reporter genes (Chalfie et al, 1994; Marshall et al, 1995; Albano, 1998; Lo et al, 1998). Because GFP fluorescence does not require any cofactors or substrates, complicated assay procedures are avoided. In addition, the protein is very stable, creating a lasting signal for detection, and background signal is low. GFP can be visualised without fixing or permeabilising cells, thus expression can be studied in live cells, along a time-course, or with a change in conditions (Albano, 1998; Persons et al, 1998). One report provides evidence that GFP might evoke an immune response (Stripecke et al, 1999). However, transgenic animals expressing GFP are viable, suggesting that this reporter gene is suitable for *in vivo* studies (Chiocchetti et al, 1997).

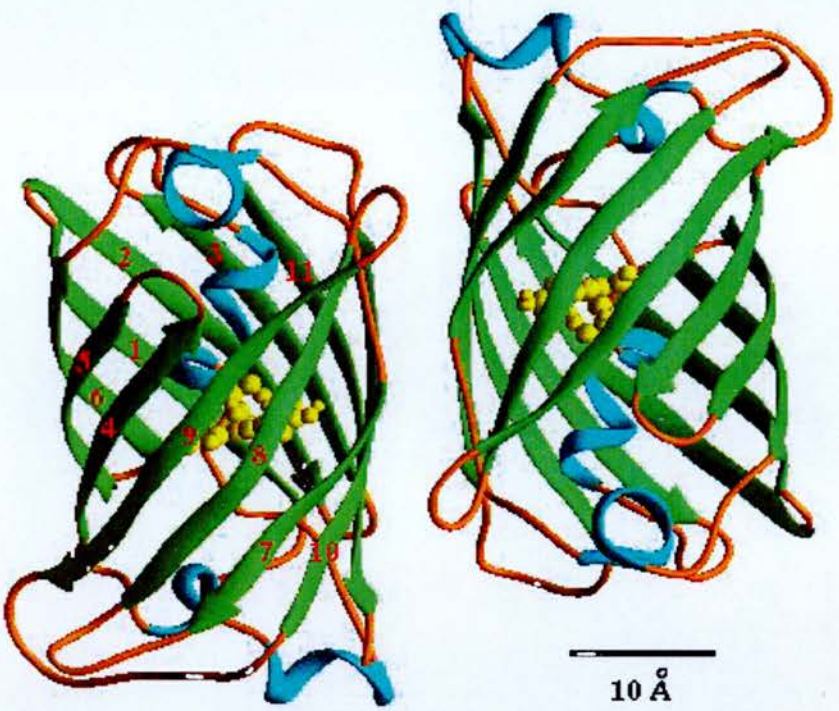


Figure 1.6. The Structure of Green Fluorescent Protein.
Eleven β-sheets (green) form the outside of the cylinder, with α-helices (blue) on the inside, in a 'β-can' arrangement. The fluorophore (yellow) lies at the center of the cylinder. This diagram was adapted from Yang et al, 1996.

Generation of enhanced GFP mutants, with increased fluorescence intensity, and mutants with shifted emissions spectra (to make red, yellow, and blue fluorescent proteins) have expanded the applications of this reporter gene (Anderson et al, 1996; Stauber et al, 1998; Clontech website). Several reports have demonstrated N- and C-terminal protein fusions with GFP, without compromising the spectral properties of GFP, or the localisation or functional properties of the fused protein (Kain et al, 1995; Marshall et al, 1995; Pines, 1995; Lo et al, 1998).

Alternatively, transcriptional fusions may be created where the genes are separated by an internal ribososme entry site (IRES), to allow separate translation of proteins from a single mRNA. IRES function was originally identified in picornvirus RNA translation (Jackson et al, 1990). The long length and predicted secondary structure of the picornvirus 5' UTR appears to be incompatible with the traditional ribosome scanning mechanism (where the ribosome binds at the 5' end of the mRNA). Thus, an alternative method must be employed to initiate translation. A 450-bp sequence from picornvirus 5' UTR was shown to confer IRES function, allowing ribosomes to bind to an internal RNA segment to initiate translation.

This IRES has practical uses in molecular biology. For example, Levenson and colleagues (Levenson et al, 1998) successfully used an IRES to drive GFP expression. They created a dicistronic construct encoding a regulated gene cassette followed by an IRES and then GFP. As the two genes were transcribed in a single mRNA before being translated separately, they were able to use GFP fluorescence to accurately model the expression pattern of the upstream gene. This would not have been feasible if these genes had been included in separate operons.

The Clontech promoterless enhanced green fluorescent protein (EGFP) vector pEGFP-1 was designed to test regulatory elements by cloning them into the multiple cloning site (MCS) upstream of the EGFP reporter gene. This approach will allow us to analyse the regulatory effects of CFTR 5' sequences. In addition, the pIRES2EGFP vector will allow the assembly of transcriptional fusions with CFTR, allowing the inclusion of intronic regulatory elements, in their natural context (Clontech website).

Several other reports strengthen the rationale for using GFP as a reporter gene in studying CFTR promoter elements. Dunn and Handelsman (Dunn and Handelsman, 1999) demonstrated the use of a promoter-trapping EGFP reporter plasmid, to identify regulatory regions. Kolossov and colleagues (Kolossov et al, 1998) demonstrated tissue-specific expression of GFP, driven by the chicken alpha-actin and beta-actin promoters. This confirms that it is appropriate to use the GFP reporter to study the effects of regulatory elements. Marshall demonstrated co-transfection of a GFP-expressing plasmid with a plasmid encoding an ion channel (Marshall et al, 1995). This facilitated identification of transfected cells; GFP did not seem to alter the properties of the co-expressed ion channels. Moyer and colleagues (Moyer et al, 1998) transfected cells with a vector encoding a GFP-CFTR fusion protein: GFP retained its fluorescent properties and CFTR appeared to retain its ion transport properties. Thus, transfection with a GFP-expressing vector in itself should not alter the ion transport properties of transfected cells.

The suitability of GFP vectors for CF gene therapy studies must be assessed. One report (Bartlett and Samulski, 1998) discussed the use of fluorescently labelled vectors for gene therapy, to identify the types of cells that are being transfected. A GFP-expressing vector under the control of a CFTR promoter could go even further in demonstrating whether the correct cell types can be targeted and signaled to express a transgene following administration to the lungs.

1.3.2 FACS analysis

Fluorescence activated cell sorting (FACS) is a powerful tool for analsying expression of the GFP gene. In this method, cells are passed through a cytometer. As each cell enters the chamber, a laser is shone upon it, and the forward scatter (FSC) – a measure of cell size, side or orthogonal scatter (SSC) – a measure of cell complexity or granularity, and fluorescent emissions FL1-H (530-560 nm; green fluorescence), FL2-H (570-610 nm; orange fluorescence) and FL3-H (610 nm +; red fluorescence) are measured (Stanford 5-minute guide to Flow cytometry website; Cancer Research UK FACS laboratory website). The data for all cells may then be assembled to produce histograms or dot-plots based on the FSC, SSC and FL-H parameters. Thus, a detailed profile is generated for the sample population.

Traditionally, histogram analysis has been used to dissect the proportion of cells expressing GFP, and the average fluorescence intensity of the transfected population, which is a measure of the level of GFP expression (this is discussed this in greater detail in Chapter 3). FACS has also been used to detect fluorescent antibody staining.

1.4 *Project aim: Towards gene therapy for cystic fibrosis: Enhanced Green Fluorescent protein as a reporter of CFTR promoter activity.*

The aim of the project was to analyse the suitability of Green Fluorescent Protein as a reporter of CFTR promoter activity. The enhanced Green Fluorescent Protein (EGFP) version was selected: this GFP variant produces brighter fluorescence than the original GFP. EGFP corresponds to the GFPmut1variant, which contains a double-AA substitution (Phe → Leu at residue 64, Ser → Thr at residue 65), as well as silent basepair changes that optimise EGFP for human codon usage (Cormack et al, 1997; Clontech website). EGFP is ~35-fold brighter than wtGFP. This choice was primarily dictated by the availability of the Clontech pEGFP-1 and pIRES2EGFP plasmids, which allow easy assembly of reporter constructs (Clontech website).

Towards this end, a series of constructs have been assembled, incorporating proximal or distal CFTR 5' regions, and/or the CFTR intron 1 region, and the EGFP reporter gene. A foldout diagram has been included in Appendix E, for easy reference while reading this thesis. A full description of these vectors is given in Chapter 3. Briefly, p1kbcfproEGFP contains 797 bp of the region 5' of the CFTR gene, driving EGFP reporter gene expression; $_{PAC}$65kbcfproEGFP contains 65 kb of CFTR 5' region driving EGFP expression: these vectors were created to allow us to assess the effects of proximal and distal 5' regulatory elements upon expression. The $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP vectors both contain 65 kb of 5' region and the coding sequence of CFTR, followed by an IRES and then EGFP. The latter contains CFTR intron 1 whilst the former does not; this allows us to assess the effects of intron 1 upon expression. The vectors pCFTRiresEGFP and $_{PAC}$RC2cmvEGFP were created as controls. These vectors allowed us to investigate the ways in which initiating transcription from an IRES, or including the reporter gene in a large PAC vector will influence expression.

The small plasmids were made by conventional cloning, while the large PAC vectors were created by a double recombination method (Boyd et al, 1999b, described in Chapter 3) incorporating both Cre/loxP and homologous recombination. Cre is a 38 kDa protein, derived from bacteriophage P1, which catalyses site-specific recombination between two loxP sites. This recombination may be intramolecular, e.g. between two loxP sites within a single DNA molecule or intermolecular, e.g. between two loxP sites on separate DNA molecules (the latter is employed in the double recombination method). The loxP site is comprised of two 13-bp inverted repeats (ATAACTTCGTATA), separated by an 8-bp asymmetric spacer region (ATGTATGC), which determines the directionality of the loxP site. Two molecules of Cre bind at each loxP site (one per inverted repeat), these catalyse a recombination event in the 8-bp spacer region (Sauer, 1993).

In order to facilitate transfection studies, several transfection reagents were compared. The liposome DOTAP has met with previous success during *in vitro* and *in vivo* studies (McLachlan et al, 1995 and 1996) and in a clinical trial (Porteous et al, 1997). The Saint-Mix™ reagent, which combines synthetic amphiphile with DOPE, has been purported to demonstrate low cytotoxicity and a high efficiency of gene transfer (Saint B.V. website). A Polyethylenimine (PEI) method condenses DNA for efficient delivery to cells (Personal communication from Matt Cotton; Thorpe et al, 2002). The LID method, described by Dr. Steve Hart, Institute of Child Health, London incorporates the liposome Lipofectin and the integrin-targeting peptide P6 (Hart et al, 1995). Finally, a new method was developed for this study; the 'SID' method combines the Saint-Mix™ reagent with the integrin-targeting peptide used in the LID method. Hence, a panel of very different transfection reagents was compared to optimise transfection efficiency in the COS-7, MDCK-IOWA, Caco-2, T84, and HBE cell lines (this is discussed in Chapter 4). EGFP expression was detected by FACS: a dotplot analysis method was found to be superior to the traditional histogram analysis method for the needs of this study.

Following these optimisation studies, the LID transfection reagent was used to transfect EGFP vectors into permanent cell lines, in order to dissect the effects of the various CFTR genomic context elements upon expression (see Chapter 5). The large $_{PAC}$RC2cmvEGFP vector was compared to the small pEGFP-N plasmid, to dissect the effects of vector size upon expression. This led to an investigation of the effects of vector copy number and plasmid mixing upon expression, with some surprising results.

Cell lines form a good starting point for gene therapy studies but are not an ideal model for CF gene therapy. Thus an *ex vivo* culture system was chosen as a more realistic model for these studies (Chapter 6). Primary sheep tracheal cells were cultured at air-interface. These cultures were characterised by transepithelial resistance, and scanning and transmission electron microscopy, and demonstrated features of the native trachea. The SID and LID methods were compared for transfecting these predominantly epithelial cultures. Finally, some of the EGFP vectors were transfected into these primary culture using the LID method.

This thesis will demonstrate that EGFP is an excellent reporter gene for studying CFTR promoter elements. The ability to detect EGFP with FACS opens new doors in analysing expression. With FACS it is possible to dissect the number of cells expressing EGFP and quantitatively measure the average fluorescence intensity of this population, which appears to be a linear measure of the level of EGFP expression (Subramanian and Srienc, 1996). A dual dotplot analysis, based on comparing a cell's FL1-H and FL2-H fluorescence, increases FACS sensitivity: allowing detection of a single fluorescent cell in a sample of 10,000 cells, or greater. This is superior to detection of the βgalactosidase reporter gene: protein assays measure the amount of βgalactosidase in a cell lysate sample, which does not distinguish between the numbers of cells expressing and the expression levels per cell, while staining assays show only the number of cells expressing the transgene but do not give a quantitative measure of the levels of expression. Other reporter genes are limited by assays with similar weaknesses. In addition, protein and staining assays for reporter genes such as βgalactosidase may be frustrated by a high background, while the FACS dual dotplot analysis provides a fairly robust distinction

of transfected cells expressing all but the lowest levels of EGFP. This is particularly useful for studies of weak CFTR promoter activity, where expression levels are likely to be low. In sum, EGFP has shown unique strengths as a reporter gene in these transfection studies and will be a suitable reporter for future experiments studying CFTR promoter elements and the mechanics of gene transfer.

Chapter 2: Materials and Methods

2.1 General methods

2.1.1 Handling PAC DNA

PAC DNA was always pipetted through cut-off tips (pipettman tips which had been widened by cutting approximately 0.5 cm off the end with a scalpel). PAC DNA was never vortexed or heated above 50 °C as this could shear or denature the delicate PAC molecule.

2.1.2 Measuring DNA concentration on a spectrometer

1 µl of the DNA sample was diluted with 99 µl of water, and placed in a Pharmica GeneQuant 100-µl cuvette. The OD_{260} was measured on a Pharmica GeneQuant RNA/DNA calculator spectrometer. The DNA concentration of the original (undiluted) sample was calculated with the formula:

Concentration in mg/ml = 5000 x OD_{260}

The DNA concentration was roughly confirmed by electrophoresis down an agarose gel, next to known masses of DNA.

2.1.3 Restriction enzyme reaction

1 µg of DNA was mixed with 2 µl of the appropriate 10X buffer (provided by Roche or NEB, with the restriction enzyme), 0.2 µl of BSA, if required, and water, to make a final volume of 19 µl, in a 1.5-ml microcentrifuge tube (Sarstedt Ltd. cat# 72.690). The tube was flicked to mix the contents, and the liquid was spun down in a picocentrifuge. 1 µl of the appropriate restriction enzyme (Roche or NEB; containing 3 to 10 U/µl, depending on the enzyme) was then added to the DNA/buffer mixture. The enzyme was aspirated and dispensed gently just under the liquid surface to avoid denaturing the delicate protein. The tube was gently flicked to mix the contents and the liquid was spun down in a picocentrifuge. The mixture was incubated at 37 °C for 24 hours. The ingredients were scaled up to digest larger masses of DNA.

2.1.4 Phosphatase

Digested DNA was treated with HK Phosphatase enzyme (1U/µl, Cambio, cat #H92050) according to the manufacturers instructions.

2.1.5 Ligation

Linearised plasmid was mixed with the insert fragment in the desired ratio, in a 1.5-ml microcentrifuge tube. 1 µl of 10X T4 ligase buffer (Roche) was added, and water was added to make a final volume of 9 µl. The tube was flicked to mix the contents, and the liquid was spun down in a picocentrifuge. 1 µl of T4 ligase enzyme (1U/µl, Roche cat#716 359) was added to this mixture; the enzyme was aspirated and dispensed gently just under the liquid surface to avoid denaturing the delicate protein. The mixture was incubated in a 4 °C water bath for 24 hours.

2.1.6 Cre/LoxP reaction

PAC DNA was mixed with the insert fragment in the desired ratio, in a 1.5-ml microcentrifuge tube. 1 µl of buffer M (Roche, see Recipes) was added, and then water was added to make a final volume of 9.5 µl. The tube was flicked to mix the contents, and the liquid was spun down in a picocentrifuge. 0.5 µl of Cre enzyme was added to this mixture. The enzyme was aspirated and dispensed gently just under the liquid surface to avoid denaturing the delicate protein. The mixture was incubated in a 30 °C water bath for one hour to allow recombination to occur.

2.1.7 Disk dialysis

A petri dish was filled with distilled water. A 0.025 µm, 25 mm white nitrocellulose filter (Millipore, cat #VSWP02500) was floated on the water. 1-5 µl aliquots of DNA were gently dispensed onto the membrane. Dialysis was allowed to proceed for 5-15 minutes; the DNA was then gently removed from the membrane with a pipettman.

2.1.8 Pouring agarose gels for electrophoresis

The required mass of agarose (Agarose type II, medium EEO, Sigma, cat # A-6877; or LMP agarose, GIBCO BRL, cat # 15517-022) was mixed with 50 ml of TBE or TAE buffer (see Recipes) in a 200-ml conical flask. The solution was microwaved at 50% power until the agar had completely dissolved. The flask containing the agar solution was cooled in a 50 °C water bath for 10 minutes. 2 µl of ethidium bromide solution (10mg/ml BioRad 161-0433) was added to the agar, if required, and swirled

to mix. The gel was then poured into a tray (tape was placed over the top and bottom edges of the tray, to contain the gel), and a comb was inserted; the gel was allowed to set at 4 °C, to expedite the cooling process. Once the gel had set, the tape was removed from the top and bottom edges of the tray (to allow current to contact the gel) and the comb was gently removed to produce sample wells. The mixture was scaled up when necessary, to make larger gels.

2.1.9 Electrophoresis

Drops of Ficoll loading buffer (approximately 1 µl of buffer per 5 µl sample volume) were dispensed onto parafilm. Each DNA sample was applied to a drop of loading buffer, and pipetted up and down to mix. The sample was then carefully loaded into a well in the agarose gel. DNA size markers were loaded alongside the samples, to provide a size reference.

The gel was placed into an electrophoresis tank containing enough buffer (TBE or TAE, corresponding to the buffer in the gel) to just cover the gel. Positive and negative electrodes were connected to a power pack (negative at the top, next to the sample wells in the agarose gel, and positive at the bottom), and a current of between 20 and 100 V was applied to the apparatus, until the DNA had run a sufficient distance down the gel.

2.1.10 Ethidium bromide staining

The agarose gel was placed in a shaking bath containing ethidium bromide solution (2 µl of Bio-Rad 10mg/ml ethidium bromide stock in 50 ml of water). The gel was stained at room temperature for 15 minutes, with gentle agitation. The gel was then transferred to a shaking bath containing water, and de-stained for five minutes, with gentle agitation, before transferring the gel to the UV transilluminator.

2.1.11 Southern Blot

PCR products were electrophoresed on an agarose gel and stained with ethidium bromide, as described above. The migrated DNA bands were visualised under the UV transilluminator, to document migration of the size markers for later reference.

The gel was placed in a shaking bath of Denature solution (see Recipes) for 15 minutes, then in a shaking bath of Neutraliser solution (see Recipes) for 15 minutes. The gel was briefly washed in water, then 2X SSC solution. Subsequently, the gel was placed on top of two pieces of thick blotting paper on a glass plate suspended above a tank. The edges of the lower piece of paper dipped into the 20X SSC in the tank below, to act as a wick. A piece of Hybond-like filter (React Scientific, cat #NOOHYA0010) was cut to the size of the gel, dipped in 2X SSC and placed on top of the gel (care was taken to remove any bubbles). Three pieces of thin blotting paper (cut to size, also dipped in 2X SSC), and then a thick layer of dry paper towels followed this. Finally, a second glass plate was placed on top and two 500-ml bottles full of water were used to weigh this down. Fig. 2.1 shows a diagram of the Southern Blot apparatus.

The apparatus was left overnight to allow the DNA in the gel to migrate through the resulting SSC gradient into the Hybond-like filter. The filter was removed and allowed to air-dry, then auto-cross linked in a Stratagene UVStratalinker 2400.

2.1.12 Hybridisation with ^{32}P-ATP labelled oligonucleotides

All radioactive work was performed in a designated area; monitoring and disposal of radioactivity was performed according to the guidelines set out by the UK National Radiological Protection Board.

Hybond-like filters from a Southern Blot were placed between pieces of gauze in a hybridisation bottle. 15 ml of Hybridisation mix (see Recipes) was added, and the bottle was pre-warmed in a rotating oven to the desired temperature for hybridisation (usually 5°C below the oligonucleotide annealing temperature).

Polynucleotide Kinase (PNK) (Roche, cat #174645) was used to label oligonucleotides with ^{32}P-ATP. 20-50 ng of oligonucleotide was mixed with 2 µl of PNK buffer, 1 µl of PNK (10U/µl) and water to make a final volume of 20 µl, in a 1.5-ml microcentrifuge tube. 1-3 µl (10-30 µCi) of ^{32}P-ATP was added to this mixture at a designated radiation workstation. The mixture was incubated at 37°C for 30-40 minutes and then added to the bottle containing the Hybaid filters. The

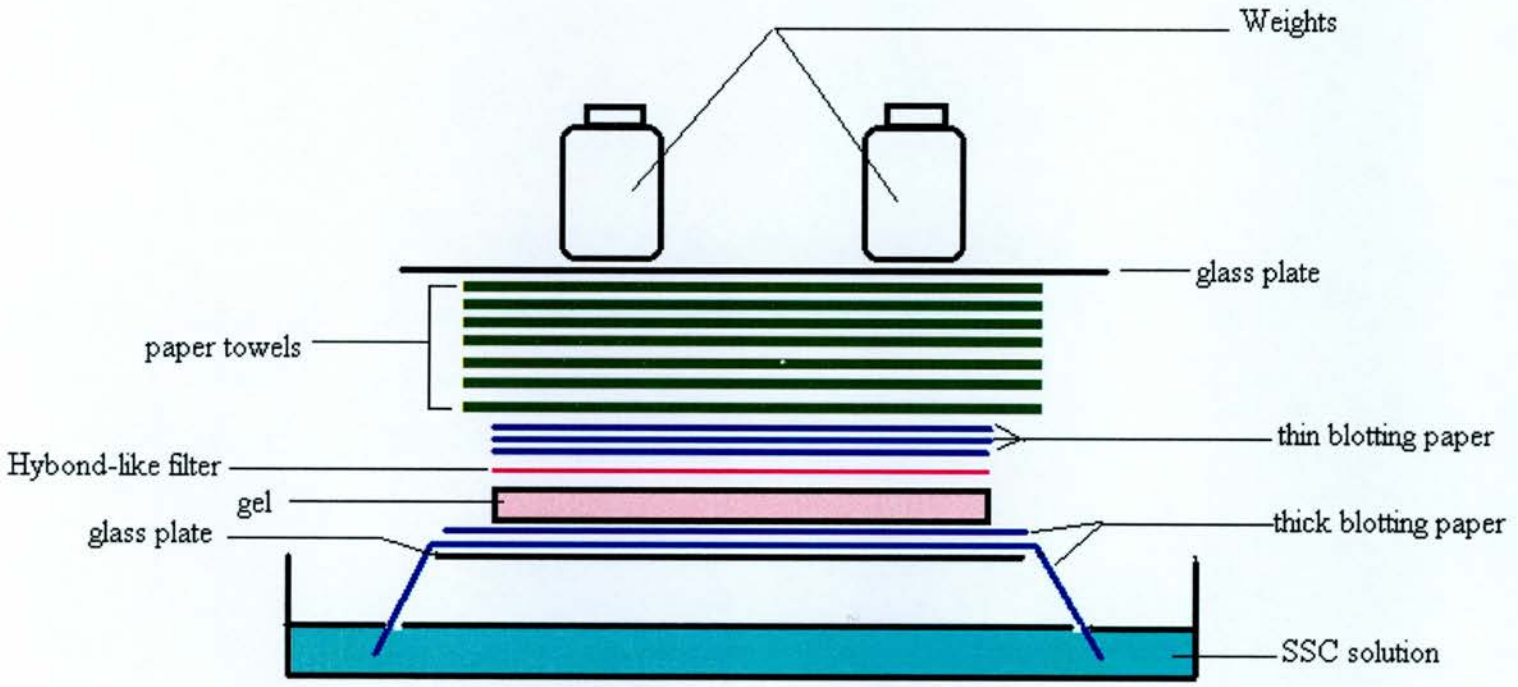


Figure 2.1. Diagram of the Southern Blot apparatus. This diagram shows the set-up of the Southern Blot apparatus. The arrangement is described in section 2.1.11.

bottle was returned to the rotating oven, and incubated for approximately 24 hours to allow hybridisation to occur.

The following day, the hybridisation mix (containing oligonucleotides that had not hybridised to the filters) was poured down the sink. The filters were washed in a 0.1% SDS, 2X SSC solution for 5 minutes. Subsequent washes were performed with solutions containing decreasing concentrations of SSC, until distinct signals of radioactivity could be discriminated from background on the filters, as determined by a Geiger counter.

The filters were placed between layers of saran wrap in a lead-lined X-ray cassette for exposure of X-ray films, or in a lead-lined phosphoimager cassette.

2.1.13 Hybridisation with ^{32}P-CTP labelled PCR products

Nitrocellulose filters from colony lifts were placed between pieces of gauze in a hybridisation bottle and 15 ml of Hybridisation mix (see Recipes) was added. The bottle was pre-warmed in a rotating oven at 68 °C.

50-100 ng of PCR product was mixed with water to make a final volume of 11 µl, in a 1.5-ml microcentrifuge tube. A hole was made in the top of the microcentrifuge tube with a needle, to prevent the tube from opening due to increasing pressure during the subsequent heating procedures. The tube was placed in a hot block at 100°C for 10 minutes to denature the PCR product.

3 µl (30 µCi) of ^{32}P-CTP and 4 µl of High Prime (Roche, cat #1585592) were added to the tube containing the PCR product, at a designated radiation workstation. The tube was then transferred to a 37 °C water bath and incubated for 20 minutes to allow High Prime labelling of the PCR product.

To ensure that the PCR product had been labelled, approximately 0.5 µl of the mixture was applied to a Whatmann GF/B 25mm filter (SLS, cat #FIL4204). The filter was then monitored with a Geiger counter, washed with 5% trichloroacetic acid (TCA), and then monitored again: if >50% of the radioactive signal remained after the wash, this indicated successful High Prime labelling.

Next, 100 µl of TNE (see Recipes) was added to the radioactively-labelled PCR product. The mixture was transferred to a hot block at 100 °C, to denature the PCR product, for 10 minutes. Subsequently, the mixture was added to the bottle containing the nitrocellulose filters, and incubated in a rotating oven at 68 °C overnight, to allow the PCR product to hybridise to the nitrocellulose filters.

The following day, the hybridisation mix (containing PCR product that had not hybridised to the filter) was poured down the sink. The filters were washed in a 0.1% SDS, 2X SSC solution for 5 minutes. Subsequent washes were performed with solutions containing decreasing concentrations of SSC, until distinct signals of radioactivity could be discriminated from background on the filters, as determined by a Geiger counter.

The filters were placed between layers of saran wrap in a lead-lined X-ray cassette for exposure of X-ray films, or in a lead-lined phosphoimager cassette.

2.1.14 GELase digestion of agarose gel slices

DNA was electrophoresed down an LMP gel (made with TAE buffer and containing ethidium bromide). Slices containing the desired DNA bands were excised from the gel under a UV transilluminator. Agarose gel slices were digested with GELase enzyme (0.2U/µl; Epicentre Technologies, cat #G31100) according to the manufacturer's instructions, with the following modifications: gel slices were melted at a temperature of 68 °C, and digestion was performed at a temperature of 42 °C.

2.1.15 Plasmid-safe DNase

Plasmid-safe ATP-dependent DNase (10U/µl; Epicentre Technologies, cat #E3101K) digestion was performed according to the manufacturer's instructions. Incubation was performed overnight at 37°C. The enzyme was either inactivated by heat denaturising (as per packet instructions) or removed by phenol-chloroform extraction (for PAC DNA only phenol-chloroform extraction was used as large PAC DNA would have been destroyed by heat inactivation). Following phenol-chloroform extraction, the DNA was ethanol precipitated to ensure the complete removal of residual phenol.

2.1.16 Polymerase chain reaction (PCR)

Reagents were mixed in the following proportions:

DNA template	1 to 100 ng
dNTPs (10 mM)	1 µl
10X Buffer	10 µl
DMSO	10 µl
Primer #1 (100 ng/µl)	1.5 µl
Primer #2 (100 ng/µl)	1.5 µl
Taq polymerase	0.2 µl
Water	to make a final volume of 50 µl

Sigma Taq DNA polymerase (5 U/µl; Sigma, cat #D-6677) was used for most reactions; the Expand™ High Fidelity PCR system (3.5 U/µl; Boehringer Mannheim, cat #1732650) was used where specified. The template was added last to prevent contamination of negative controls. When multiple reactions were being set up, the mixture was made up in bulk, and then aliquoted. Reactions were set up in either 0.5-ml microcentrifuge tubes (Sarstedt Ltd. cat# 72.735.002), or 96-well plates (96-well plate covers were washed with ethanol and water, then sterilised by 15 minutes of UV exposure in the Stratagene UVStratalinker 2400).

PCR was performed on an MJ Research PTC-225 or PTC-200 Peltier thermal cycler. The following programs were used:

'65kbcfpro' (for construction of $_{PAC}$65kbcfproEGFP)

STEP 1:	94°C – 0:10
STEP 2:	55°C – 1:00
STEP 3:	72°C – 2:00
STEP 4:	go to step 1, 2 times
STEP 5:	94°C – 0:10
STEP 6:	65°C – 1:00
STEP 7:	72°C – 2:00
STEP 8:	go to step 5, 26 times
STEP 9:	94°C – 3:00
STEP 10:	4°C for ever

'CF5491PCR'
(for construction of $_{PAC}$RC1iresEGFP, $_{PAC}$RC2iresEGFP and $_{PAC}$RC2cmvEGFP)

STEP 1:	94°C – 0:10
STEP 2:	55°C – 1:00
STEP 3:	72°C – 2:00
STEP 4:	go to step 2, 29 times
STEP 5:	94°C 3:00
STEP 6:	4°C for ever

'65PCR' (for colony PCRs)

STEP1:	95°C – 1:30
STEP 2:	95°C – 0:30
STEP 3:	65°C – 1:00
STEP 4:	72°C – 1:00
STEP 5:	go to step 2, 29 times
STEP 6:	4°C for ever

'60PCR' (for colony PCRs)

STEP1:	95°C – 1:30
STEP 2:	95°C – 0:30
STEP 3:	60°C – 1:00
STEP 4:	72°C – 1:00
STEP 5:	go to step 2, 29 times
STEP 6:	4°C for ever

'βgal' (for βgalactosidase and GAPDH RT-PCRs)

STEP 1:	94°C – 5:00
STEP 2:	95°C – 0:30
STEP 3:	50°C – 0:10
STEP 4:	72°C – 4:00
STEP 5:	go to step 2, 29 times
STEP 6:	4°C for ever

'GFP' (for GFP RT-PCRs and colony PCRs)

STEP 1:	96°C – 1:30
STEP 2:	94°C – 0:30
STEP 3:	52°C – 1:00
STEP 4:	72°C – 1:00
STEP 5:	go to step 2, 29 times
STEP 6:	4°C for ever

2.1.17 Hot-start PCR

The hot-start PCR protocol was designed to enhance amplification from small amounts of DNA (Chou et al, 1992). Heating the template DNA and primers separately, then combining them for PCR while they are still hot reduces mispriming and primer dimersiation. This protocol is useful to amplify DNA from large PAC templates where the possibility of mispriming may be enhanced, due to the presence of sequences similar to the primer-annealing site in this larger template molecule. This protocol may also be useful when using large primers, or primers with a predicted secondary structure (e.g. the loxP primer), to prevent annealing between or within primers.

For hot-start PCR, a reaction mixture was made up as described above, omitting the template DNA, and reducing the amount of water, to make a final volume of 25 µl per reaction. This mixture was made up in bulk in a 0.5-ml microcentrifuge tube. For each reaction, 10-100 ng template DNA was mixed with water to make a final volume of 25 µl per reaction in a 0.5-ml microcentrifuge tube. The individual tubes of template DNA (one per reaction) and tubes containing the bulk reaction mixture were placed in the PCR block, and the following program was run:

'HOTSTART' (for construction of p1kbcfproEGFP)

STEP 1:	94°C – 3:30*
STEP 2:	94°C – 0:10
STEP 3:	55°C – 1:00
STEP 4:	72°C – 2:00
STEP 5:	go to step 2, 2 times
STEP 6:	94°C – 0:10
STEP 7:	65°C – 1:00
STEP 8:	72°C – 2:00
STEP 9:	go to step 6, 26 times
STEP 10:	94°C – 3:00
STEP 11:	4°C for ever

*After two minutes at 94°C (during the first step), the tubes in the PCR block were opened, and 25 µl of the bulk reaction mixture was added to each tube containing template. This procedure was completed quickly, before the first step had ended. The tubes were then closed, and the PCR program was allowed to run to completion.

2.1.18 Colony PCR

For colony PCRs, the PCR reaction mixture was inoculated with bacteria from a single colony or streak, using a toothpick (in place of the template DNA). Otherwise, colony PCRs were performed with the methods and programs described above.

2.1.19 DEPC treatment of water

100 µl of DEPC (SIGMA, cat #D5758) was added to 1L of UHP dH_20 water. This solution was mixed, and left in a fume cupboard overnight to allow DEPC to break down any RNase present in the water. The solution was autoclaved before use, to break down the remaining DEPC.

2.1.20 RNA extraction

RNAzol cell lysates were collected in 1.5-ml screw-top microcentrifuge tubes (see Tissue Culture Methods). Samples were prepared in an Envair class II microbiological safety cabinet, to prevent contamination of the prep with DNA or RNase.

100 µl of chloroform was added to each sample. The mixture was vortexed for 10 seconds, and incubated on ice for 5 minutes. The aqueous and organic phases were separated by spinning in a refrigerated microcentrifuge at 13,000 RPM for 15 minutes. 450 µl of the top (aqueous) phase was transferred to a new tube.

To precipitate the RNA, 450 µl of isopropanol and 4 µl of glycogen (to act as a carrier) were added to each tube. Samples were left on ice for 15 minutes to precipitate the RNA, and then spun in a refrigerated microcentrifuge for 30 minutes at 13,000 RPM. The supernatant was removed and the pellet was washed by adding 1 ml 75% ethanol and vortexing for 10 seconds. RNA was pelleted as before in a refrigerated microcentrifuge for 15 minutes.

The supernatant was completely removed and the RNA pellet was re-suspended in 49 µl of DEPC-treated water.

2.1.21 DNase Treatment

Samples were prepared in an Envair class II microbiological safety cabinet, to prevent contamination of the prep with DNA or RNase.

RNA samples were prepared in a volume of 49 µl. The following reagents were then added:

DNase 1 (RNase free) 10 U/µl (Roche)	2 µl
RNaseIn 50 U/µl (Roche, cat #799017)	1 µl
10X DNase buffer	5.8 µl

Samples were mixed by flicking and then spun down for 30 seconds in a microcentrifuge. The mixture was incubated at 37 °C for 60 minutes.

Phenol-chloroform extraction was performed to remove residual enzymes: 60 µl of a 1:1 mixture of phenol: chloroform was added to each tube and the samples were vortexed for 10 seconds. The samples were spun in a refrigerated microcentrifuge at 13, 000 RPM for 15 minutes, to separate the organic and aqueous phases, and approximately 55 µl of the top (aqueous) phase was transferred to a new tube on ice.

55 µl of chloroform was then added (to remove any residual phenol); the samples were again vortexed and spun for an additional 15 minutes. Approximately 50 µl of the top (aqueous) layer was transferred to a new tube.

5 µl of 3M NaOAc (pH 5.6) and 120 µl of ethanol were added to each sample. Samples were transferred to dry ice for 15 minutes (or placed in a -70 °C freezer overnight) to precipitate the RNA. RNA was pelleted in a refrigerated microcentrifuge for 30 minutes at 13,000 RPM. Samples were transferred immediately to ice.

The supernatant was removed and the pellet was washed with 1 ml of 75% ethanol and vortexing. RNA was pelleted by an additional 15-minute spin in the microcentrifuge. The supernatant was completely removed and the pellet was re-suspended in 20 to 50 µl of DEPC-treated water.

RNA was stored in a -70 °C freezer before cDNA synthesis, if necessary.

2.1.22 cDNA synthesis

A first strand cDNA synthesis kit (Roche, cat #1483188) was used.

Samples were prepared in an Envair class II microbiological safety cabinet, to prevent contamination of the prep with DNA or RNase.

The RNA samples were thawed on ice if necessary. 2-5 µl of each RNA sample was aliquoted into a 1.5-ml screw top microcentrifuge tube (duplicates were included for each sample to provide a control without AMV-RT – see below). DEPC-treated water was added to make a total volume of 7.8 µl (negative controls of water only were also included). These samples were denatured at 65 °C for 15 minutes and then chilled on ice.

A bulk reaction mixture was prepared containing the following quantities of reagents for each sample:

10X buffer	2 µl	(Tube 1)
$MgCl_2$	4 µl	(Tube 2)
dNTPs	2 µl	(Tube 3)
Gelatine	0.4 µl	(Tube 4)
$P(dN)_6$ primer	2 µl	(Tube 6)
RNaseIn	1 µl	(Tube 7)
AMV -RT	0.8 µl	(Tube 8)

A duplicate mixture was made up with water in place of the AMV-RT, to ensure that any signal in the subsequent RT-PCR was due to cDNA created from the RNA rather than a background signal from genomic/plasmid DNA (which is a common artefact due to incomplete DNase I digestion).

The heat-denatured RNA was spun down for 30 seconds. 12.2 µl of the mixture containing AMV -RT was added to each RNA sample, and 12.2 µl of the mixture without AMV -RT was added to the duplicate RNA samples. The mixture was pipetted up and down several times to mix.

The samples were incubated at room temperature for 10 minutes to allow the primers to anneal, then at 42 °C for 60 minutes.

The sample was spun down and kept on ice while preparing for RT-PCR, or frozen at -70 °C.

2.1.23 Reverse transcription polymerase chain reaction (RT-PCR)

Samples were prepared in an Envair class II microbiological safety cabinet, to prevent contamination of the prep with DNA.

5 µl of each cDNA sample was aliquoted into a labelled 0.5-ml microcentrifuge tube. Samples + AMV-RT and –AMV-RT were included for each RNA sample, in addition, negative controls (containing water in place of RNA) and positive controls (containing plasmid DNA) were included.

A bulk mixture was prepared containing the following quantities of reagents for each sample:

10 X buffer (Roche):	5 µl
DMSO	5 µl
10 mM dNTPs	1µl
Primer 1 (100 ng/µl)	1 µl
Primer 2 (100 ng/µl)	1 µl
Sigma Taq polymerase 5 U/µl	0.5 µl
DEPC water	31.5 µl

45 µl of this bulk mixture was added to each sample, and mixed by pipetting.

PCR was performed on an MJ Research PTC-225 or PTC-200 Peltier thermal cycler, as described above.

2.1.24 ExoSAP-IT™

5 µl (containing approximately 50 ng) of PCR product was mixed with 1 µl of ExoSAP-IT™ (USB, cat #78201) in a 96-well plate. The plate was heated on an MJ Research PTC-225 Peltier Thermal Cycler at 37°C for one hour, and then the enzyme was denatured at 80°C for 20 minutes.

2.1.25 Qiaquick PCR purification

Qiaquick PCR purification was performed according to the kit protocol. DNA was eluted from the column in a 30-µl volume, to increase DNA concentration in the eluate. Water was used to elute the DNA, as salts in the kit elution buffer might have interfered with the subsequent sequencing procedure.

2.1.26 Sequencing

6 µl (containing approximately 10ng) of template DNA was combined with 6 µl of primer (10 ng/µl), 4 µl of Rhodamine reaction mixture, and 4 µl of half-term buffer in a 96-well plate. Cycle-sequencing was performed on an MJ Research PTC-225 Peltier Thermal Cycler with the following program:

STEP1:	95 °C, 30 seconds
STEP2:	53 °C, 20 seconds
STEP3:	60 °C, 3 minutes
GOTO	step 1, 40 X

2.1.27 PCR sequencing

PCR sequencing was performed as above, but 6 µl of ExoSapIT™ reaction mixture (containing PCR product) was used in place of the template DNA.

2.1.28 Cleaning up sequencing reactions

Reagents were mixed in the following proportions:

95% ethanol (sequencing grade)	50 µl
3M sodium acetate (NaOAc), pH 4.6	2 µl
Pellet Paint co-precipitant (Novagen, cat# 69049)	0.5 µl

A bulk mixture was made up and 52.5 µl of this was added to each 20-µl sequencing reaction in a 0.5-ml microcentrifuge tube. The mixture was vortexed for 10 seconds and incubated at room temperature for one hour to precipitate extension products.

The extension products were pelleted in a microcentrifuge for 30 minutes at 13,000 RPM. The supernatant was removed and the pellet was washed with 250 µl of 70 % ethanol. The tube was vortexed to fully wash the pellet and centrifuged for 15 minutes at 13,000 RPM in a microcentrifuge. The washing step was then repeated.

Following this, the supernatant was completely removed and the pellet was allowed to air-dry for 1 minute. Pellets were stored at 4 °C until rehydration and electrophoresis. The sequences were run on an ABI Prism DNA sequencher as a service by Angie Fawkes in the Clinical Genetics Section, MMC University of Edinburgh.

The sequencing results were viewed, and a Contig was created from the sequences for each clone, using the 'Consed' software package. The Human Genome Mapping Project (HGMP) GCG software 'Bestfit' was used to compare this Contig sequence to the expected sequence.

2.1.29 Ethanol precipitation

DNA was mixed with one half-volume of 3M Sodium Acetate (NaOAC) pH 4.6 and four new volumes of 100 % ethanol. The mixture was vortexed to mix (for PAC DNA the mixture was flicked to mix, as vortexing would have sheared the large PAC DNA molecule). The mixture was then spun for 30 minutes at 13,000 RPM in a microcentrifuge to pellet the DNA. The supernatant was removed and the pellet was washed in 70% ethanol and vortexed or flicked to mix. The pellet was then re-centrifuged at 13,000 RPM for 15 minutes. The supernatant was removed completely and the pellet was allowed to air-dry for one minute. DNA was then re-suspended in a suitable volume of 1 X TE (pH 8.0) or water.

2.1.30 Phenol: chloroform extraction

An equal volume of PCI (see Recipes) was added to the aqueous solution in a 1.5-ml microcentrifuge tube. The mixture was vortexed or inverted to mix. The mixture was spun at 13,000 RPM in a microcentrifuge for one minute, to separate the organic and aqueous phases. The top (aqueous) phase was removed to a fresh 1.5-ml microcentrifuge tube; care was taken to avoid debris at the phase interface. If necessary, a chloroform extraction was then performed to remove any traces of phenol (e.g. the procedure was repeated, using neat chloroform in place of PCI).

2.2 Bacterial Methods

2.2.1 Making competent cells

Soap-free glassware was used throughout the protocol to improve cell transformation competence.

A plate was poured containing 0.4 g of DIFCO Bacto agar (SLS, cat#0140 – 01) in 35 ml SOB –Mg (see Recipes), containing IPTG (626 µg/ml) if required. The competent cell stock was streaked out on this plate and grown in a 37°C incubator overnight to obtain single colonies.

4 X 800-ml cultures were set up containing SOB –Mg and IPTG (625 µg/ml), if required. A single colony was inoculated into each culture. These cultures were grown at 30°C in a shaking incubator for approximately 15 hours until an OD_{550} of 0.75 - 1 was reached.

16 X 50-ml Falcon tubes were filled with the contents of one 800-ml culture. The cells were spun down at 4,000 RPM for 15 minutes in a Jouan centrifuge and the supernatant was discarded. The process was repeated with the other three 800-ml cultures, pouring them into the same 16 tubes to pool the cells.

The cells were re-suspended in 50 ml of cold GYT (see Recipes) to wash, and pelleted by centrifugation. This step was repeated. The cells were then re-suspended in a small volume of GYT, and pooled into eight falcons; 50 ml of cold GYT was added to wash and the cells were again pelleted by centrifugation. This process was repeated to pool the cells into four, and then two falcons with a GYT wash between each consolidation. After the final spin, most of the GYT supernatant was poured off from the pellet, leaving a thin layer. The cell pellet was re-suspended in the remaining GYT to produce a thick suspension. 140-µl aliquots of the cell suspension were snap-frozen by dispensing into 1.5-ml microcentrifuge tubes on dry ice. Competent cell aliquots were stored at -70°C. A test transformation was performed on each batch prior to use, to confirm competence and rule out bacteriophage contamination.

2.2.2 Electroporation

Aliquots of competent cells were thawed on wet ice. 20 – 30 µl of the competent cell suspension was combined with the DNA (usually between 10 and 200 ng) in a 1.5-ml microcentrifuge tube. This mixture was flicked to mix, then transferred to a chilled 1 mm cuvette and tapped down to ensure the mixture had entered the well between the electrode contacts in the cuvette. A pulse of 1.25 V at 100 Ω, 25 µF was applied to the cells with a BioRad Gene Pulser™ electroporator. 0.5 – 1 ml of recovery medium (10 µl/ml glucose in L-broth) was immediately added to the cells, and then the mixture was transferred to a sterile 5-ml polystyrene tube with vent-cap (Greiner, cat# 120180). The cells were allowed to recover for 1 hour in a 37°C shaking incubator. Aliquots of the transformed cells were spread on L-agar plates (containing antibiotics at 50 mg/ml, IPTG at 625 µg/ml, and XGal at 625 µg/ml, if required), and incubated at 37°C overnight to allow colony growth.

When IPTG/Xgal staining was used to identify colonies expressing βgalactosidase, the plates were incubated at 4°C for two hours, to allow the blue colour to develop, prior to scoring the plates.

2.2.3 Streaking out bacteria for single colonies

An L-agar plate was prepared containing the appropriate antibiotics (50µg/ml) and IPTG/XGal (each at 625 µg/ml), if required.

A plastic loop was used to collect a sample from a bacterial colony or frozen stock. This was streaked across the L-agar plate. A new loop was used to make a second streak from the end of the first. The process was repeated to create a pattern of streaks with decreasing concentration of bacteria. The plates were incubated at 37°C overnight to allow colony growth: a portion of the plate would then contain single colonies.

2.2.4 Patching colonies in grid formation

A plate of L-agar (containing the appropriate antibiotics at 50 µg/ml and IPTG/Xgal at 625 µg/ml, if necessary) was prepared, and laid on top of a 96-square grid. The plate was marked to confirm orientation. A small sample of the bacterial colony was

collected onto a plastic loop, and streaked diagonally across one of the boxes in the grid. This process was repeated for each colony. The plate was incubated at 37°C overnight, to allow colony growth.

2.2.5 Preparing frozen stocks of transformed bacteria

A single colony was inoculated into L-broth containing the appropriate antibiotics (50 µg/ml) in a conical flask. The culture was grown overnight at 37°C in a shaking incubator. The next day, 1 ml of the culture was mixed with 1 ml freezing mix (60% SOB, 40% Glycerol) in a cryotube. The stock was stored at -70°C.

2.2.6 Crude DNA minipreps

Bacterial colonies were inoculated into 2-ml aliquots of L-broth containing the appropriate antibiotics (50 mg/ml), in 5-ml Bijoux tubes (the caps were loosened to allow air circulation). The Bijoux tubes were placed in a basket in the 37°C shaking incubator, and incubated overnight.

1.5 ml of each cell suspension was poured into a 1.5-ml microcentrifuge tube. The cells were spun down in a microcentrifuge for five minutes at 13,000 RPM. The supernatant was removed and the pellet was re-suspended in 100 µl GTE (see Recipes). This suspension was vortexed to mix, and incubated at room temperature for five minutes.

200 µl of 0.2M NaOH, 1% SDS solution was added; the tube was inverted to mix and incubated for five minutes at room temperature.

150 µl of 3M KAc was added; the tube was inverted to mix and was incubated on ice for five minutes.

The tube was then spun in a microcentrifuge at 13,000 RPM for 10 minutes to precipitate proteins. The supernatant was transferred to a fresh 1.5-ml microcentrifuge tube.

400 µl of PCI (see Recipes) was added to the sample. The mixture was vortexed and spun in a microcentrifuge at 13,000 RPM for one minute, to separate the aqueous and organic phases.

400 µl of the top (aqueous) phase was removed to a new 1.5-ml microcentrifuge tube (avoiding any debris at the interface).

1 ml of ethanol was added to the solution; the mixture was flicked to mix and spun in a microcentrifuge at 13,000 RPM for 30 minutes to precipitate DNA.

The supernatant was removed and 1 ml of 70% ethanol was added to wash the DNA pellet. The pellet was then re-centrifuged for 15 minutes.

The supernatant was removed with a pipettman, and the pellet was left to air-dry for five minutes. The pellet was then re-suspended in a small volume of TE pH 8.0, or water.

2.2.7 Qiagen Midipreps and Maxipreps

QIAGEN midipreps (Qiagen, cat # 12143) and maxipreps (Qiagen, cat #12163) were performed according to the kit instructions with the following exceptions:

After incubation with buffer P3, the solution was centrifuged only once then passed through gauze to ensure complete removal of particulate material before adding this solution to the QIAGEN-tip.

Isopropanol precipitation was performed in 50-ml Falcon tubes. The DNA pellet was precipitated by centrifugation in a Jouan centrifuge for one hour at 4,000 RPM. The DNA pellet was washed with 70% ethanol and a second spin was performed at 4,000 RPM for 30 minutes. The pellet was re-suspended in a small volume of TE pH8.0, or water.

When PAC DNA was prepared, the media volumes were altered. A larger volume of culture was set up for each QIAGEN-tip: 50 ml for each midiprep tip and 200 ml for each maxiprep tip. A larger volume of P1, P2, and P3 were used: 5 ml of each solution per 50 ml culture volume. In addition, QIAGEN-tips were washed three times with QC buffer, before the elution step. These alterations were based on an optimised protocol for the preparation of PAC DNA, developed by Ann Doherty and Heather Davidson (members of the MMC Medical Genetics group, University of Edinburgh).

A Beckman Coultier Avanti J20 I centrifuge with JA25 rotor was used to pellet the bacteria and bacterial debris, prior to applying the supernatant to the QIAGEN-tip. A Jouan CR3i centrifuge was used for DNA precipitation.

2.2.8 Colony lifts

Colonies were patched in grid formation and grown overnight. A round nitrocellulose filter (Protran BA 85/20 0.45 mm, 82 mm filter, Schleicher and Schuell) was labelled and marked for orientation. This membrane was laid on top of the agar plate containing the patched bacterial colonies. The filter was left for a few seconds, to allow the colonies to adhere.

The nitrocellulose filter was then laid colony side-up, on top of pieces of blotting paper soaked in the following solutions, for the specified periods:

10% SDS	– 5 min.
Denature	– 5 min.
Neutraliser	– 5 min.
0.4M NaOH	– 20 min.

The membrane was then rinsed twice in 2 X SSC, 0.1% SDS solution, then once in 2 X SSC solution, with gentle agitation to remove colony debris. The membrane was then cross-linked in the Stratagene UVStratalinker 2400, to permanently bind the bacterial DNA.

2.3 Tissue culture Methods

2.3.1 Culture of immortalised cell lines

Immortalised cell lines were handled in an Envair class II microbiological safety cabinet, to prevent contamination with microorganisms.

2.3.2 Splitting cells

Growth medium was removed from a flask of cells (Greiner 75 cm flasks with vented caps, cat #658175 and 25cm flasks with vented caps, cat #690175, were used). A 1:1 mixture of PBS and Versine was added to the flask (10 ml for a 75cm flask, 5 ml for a 25cm flask) to wash away remaining medium and inactivate any residual serum. This mixture was removed and an equal volume of a 1:1 mixture of

Trypsin and Versine was added to the flask. The flask was placed in a humidified 37°C, 5% CO_2 incubator for approximately 10 minutes until cells had detached from the plastic and dissociated (confirmed by visual examination under the microscope). It was often necessary to pipette the cells several times through a fine-tipped pastette to produce a single-cell suspension.

New flasks of cells were set up as follows:

Table 2.1. Seeding flasks with permanent cell lines.

Cell line	Fraction of cells from old flask used to seed new flask:
COS7	1/10
MDCK-IOWA	1/10
CaCO2	1/3
HBE	1/3
T84	1/3

The appropriate fraction of the cell suspension was transferred to a new flask and growth medium was added (30 ml for a 75cm flask, 10 ml for a 25cm flask).

Cells were grown in a humidified 37°C, 5% CO_2 incubator. The growth media was changed twice a week and the cells were split once a week.

2.3.3 Collagen-coating Transwell inserts

25 mg of human placental collagen type VI, acid soluble (Sigma, cat # C-7521) was mixed with 50 ml tissue culture tested deionised H_2O and 100 µl of glacial acetic acid. This mixture was heated in a 37°C water bath until the collagen dissolved, then an additional 450 ml H_2O was added. This solution was filtered through a 0.2 µm acrodisc prior to use.

100 µl of collagen solution was added to each Costar Transwell Clear insert (6.5mm diameter, 0.4 mm pore, Costar C3470) in a 24-well plate. This was left in the laminar hood to coat for 6 to 24 hours. Before use, any remaining liquid collagen was removed from the insert and the membrane was washed inside and out with PBS.

2.3.4 Establishing sheep primary tracheal air-interface cultures

Intact sheep tracheae were obtained from the Moredun Research Facility at Easter Bush, Edinburgh. The tracheae were transferred to PBS + 1% Penicillin /Streptomycin (Stock 10,000 u/ml, 10,000 µg/ml, GIBCO, cat # 15140-122), Amphotericin B (0.25 µg/ml) and Vancomycin (100 µg/ml) for transport to the laboratory.

(Day 1) As much fat and connective tissue were removed from the outside of the trachea as possible, and an incision was made along the length of the trachea on the posterior side, between the extremities of the cartilage rings. Muscle and membrane at the border of the cartilage rings were trimmed away. The trachea was sliced between the cartilage rings at intervals of two to three rings; these pieces were then cut into squares of roughly 1 cm^2. The pieces of trachea were added to 50 ml of Dissociation media (see Recipes) in a 50-ml falcon, and incubated overnight at 4°C, to allow the cells to dissociate.

(Day 2) The Falcon tubes were inverted gently 6X and the dissociated cell suspension was poured off the tracheal husks. The tracheal husks were washed with 50 ml of Airway media (see Recipes) to harvest additional cells, and this was combined with the first cell suspension. FCS was added to a final volume of 10%.

Cells were spun down from the suspension, and then re-suspended in 10 ml of PBS. To lyse any blood cells in the preparation, 20 ml of tissue culture treated dH_2O was added. After 30 seconds, 2 ml of 10X PBS was added to restore isotonicity.

The mixture was again spun down, and cells were re-suspended in 48 ml of Airway media per trachea. 12-ml aliquots were added to 10 cm tissue-culture plates (Primeria; Falcon cat #3803). These were incubated for 4-5 hours to remove any fibroblasts, which adhere preferentially to the plastic.

The remaining cell suspension containing predominantly epithelial cells was collected and four 12-ml aliquots were pooled into a 50-ml Falcon tube. A 100-µl aliquot of the cells was mixed with 100 µl of tryphan blue, to stain dead cells. The mixture was then applied to a haemocytometer slide and visualised under the light

microscope, to assess cell viability and quantify live cells. The cells were spun down and then re-suspended in 200 µl of Airway media per 1 X 10^6 cells. A 200-µl volume of the cell suspension was seeded on the inside of each collagen coated Costar insert and 600 µl of Airway media was placed below the insert to feed the cells from below. The cells were cultured in a 37°C humidified incubator with 6% CO_2.

(Day 3) Media below the insert was replaced with 600 µl USG media (see Recipes) and media inside the insert was removed, creating an air-interface. Cells were cultured for up to 5 weeks. The USG growth media was replaced twice a week.

Visual examination was used to evaluate the presence of ciliated cells. When focusing just above the plain of the cells, a shimmering effect could be seen above some cells in the culture. This was indicative of beating cilia (personal communication from Donald Davidson).

A System EVOM Epithelial Voltohmeter + STX2 Electrode Set (World Precision Instruments, Inc) was used to monitor electrical resistance across the membrane. Growth media below the insert was replaced with 600 µl PBS and 200 µl PBS was placed above the insert. One prong on a set of electrodes was lowered into the PBS below the insert, and the other into the PBS above the insert, and the resistance reading was taken. A high resistance reading (above 1,000 Ω/membrane) indicated the presence of tight junctions between the cells. Readings were taken every few days to assess the presence of tight junctions throughout the culture period.

2.3.5 Transfection

Permanent cell lines were seeded as follows:

Table 2.2. Seeding permanent cell lines for transfection.

Cell line	Number of cells per well (24-well plate)
COS7	5 X 10^4
MDCK-IOWA	5 X 10^4
CaCO2	1 X 10^5
HBE	1 X 10^5
T84	5 X10^5

The appropriate number of cells was added to each well in a 24-well plate. 2 ml of growth media was added to each well. The cells were transfected at 60-80% confluence (this density was usually achieved 24 hours after seeding). Cells were washed once with 2 ml of serum-free growth medium, prior to transfection.

2.3.5.1 Dotap Transfection

12 µl of DOTAP liposomal transfection reagent (Boerhinger-Manheim, cat # 1202375) was mixed with 88 µl of Optimem 1 in a sterile 5-ml polystyrene tube (Falcon #352054). In a separate tube, 1 µg of DNA was mixed with Optimem 1 to make a total volume of 100 µl. The two solutions were combined and shaken vigorously to mix. The mixture was left for 20 minutes to allow a complex to form.

300 µl of Optimem 1 was added to the complex, to make a final volume of 500 µl; this 500-µl volume was applied to a single well in a 24-well plate. For duplicate and triplicate transfection, the mixture was scaled up and made in bulk.

After a five-hour transfection period, the transfection complex was removed and 2 ml of growth media was added to each well.

2.3.5.2 Saint Mix Transfection

20 µl of Saint-Mix™ (Saint B.V.) was mixed with 80 µl of Optimem 1 (GIBCO, cat #319-047) in a sterile 5-ml polystyrene tube. In a separate tube, 1 µg of DNA was mixed with Optimem 1 to make a total volume of 100 µl. The two solutions were then combined and pipetted several times to mix (with PAC DNA, the tubes were shaken in preference to pipetting, to prevent shearing of the DNA). The mixture was used immediately.

For transfection of permanent cell lines, 300 µl of Optimem 1 was added to the complex, to make a final volume of 500 µl; this 500-µl volume was applied to a single well in a 24-well plate. For primary cultures, the complex was used undiluted; the 200-µl volume was applied to a single primary culture insert (see below). For duplicate and triplicate transfections, the mixture was scaled up and made in bulk.

After a four-hour transfection period, 2 ml of growth media was added to each well without removing the complex. 24 hours later, the growth media was replaced. For primary cultures, the transfection complex was simply removed from the cell surface after four hours.

2.3.5.3 PEI transfection

3 µg of DNA was mixed with HBS to make a 125-µl final volume, in a sterile 1.5-ml microcentrifuge tube. In a separate tube, 2.5 µl of PEI (2K) was mixed with HBS to make a 125-µl final volume; this mixture was then combined with the DNA/HBS. The mixture was flicked to mix then contents, then incubated at room temperature for 15 minutes.

In a separate tube, 4.5 µl of PEI (25K) was mixed with HBS to make a 125-µl final volume. This mixture was then added to the DNA/ PEI (2K) mixture. The tube was flicked to mix the contents, and then incubated at room temperature for 15 minutes.

500 µl of serum-free media was added to each well in a 24-well plate. 120 µl of complex (containing 1 µg DNA) was added to each well, the plate was gently swirled to mix. After a four-hour incubation at 37 °C, the transfection complex was removed and 2 ml of growth media was added.

The mixture was not scaled up or down, as this appears to compromise transfection efficiency (personal communication, Matt Cotton).

2.3.5.4 LID transfection

0.75 µl of Lipofectin reagent (GIBCO BRL, cat #18297-011) was mixed with 40 µl of P6 integrin-targeting peptide (100ng/µl in Optimem 1) and 60 µl of Optimem 1 in a sterile 5-ml polystyrene tube. 1 µg of DNA was mixed with Optimem 1 to make a final volume of 100 µl. The two solutions were then combined. The mixture was used immediately.

For transfection of permanent cell lines, 300 µl of Optimem 1 was added to the complex, to make a final volume of 500 µl; this 500-µl volume was applied to a

single well in a 24-well plate. For primary cultures, the complex was used undiluted; the 200-µl volume was applied to a single primary culture insert (see below).

After a five-hour transfection period, the transfection complex was removed and 2 ml of growth media was added to each well. For primary cultures, the transfection complex was simply removed from the cell surface after five hours. For duplicate and triplicate transfections, the mixture was scaled up and made in bulk.

2.3.5.5 SID transfection

20 µl of Saint-Mix™ (Saint B.V.) was mixed with 10 - 40 µl of P6 integrin- targeting peptide (100ng/µl in Optimem 1) and Optimem 1 to make a final volume of 100 µl, in a sterile 5-ml polystyrene tube. In a separate tube, 1 µg of DNA was mixed with Optimem 1 to make a final volume of 100 µl. The two solutions were then combined and pipetted several times to mix (with PAC DNA, the tubes were shaken in preference to pipetting, to prevent shearing of the DNA). The mixture was used immediately.

For transfection of permanent cell lines, 300 µl of Optimem 1 was added to the complex, to make a final volume of 500 µl; this 500-µl volume was applied to a single well in a 24-well plate. For primary cultures, the complex was used undiluted; the 200-µl volume was applied to a single primary culture insert (see below).

After a five-hour transfection period, the transfection complex was removed and 2 ml of growth media was added to each well. For primary cultures, the transfection complex was simply removed from the cell surface after five hours. For duplicate and triplicate transfections, the mixture was scaled up and made in bulk.

2.3.6 Procedure for transfection of primary cultures

The resistance across the membrane was measured with a voltohmeter before transfection to assess the presence of tight junctions (see above); only cultures with a resistance > 1,000 were accepted for transfection. Cultures of relatively higher and lower resistances were spread evenly between transfection conditions, to prevent a bias in the results.

Tight junctions were disrupted by applying 200 µl of 6 mM EGTA (pH 7.0) above the membrane, and 600 µl below. The cultures were incubated at 37 °C for 20 minutes, then the new resistance readings were measured: a resistance < 350 Ω/membrane indicated successful disruption of tight junctions. The PBS used to measure resistance also served to wash serum away from the cells before transfection.

As described above, 200 µl of transfection mix (containing 1 µg DNA) was placed above each insert, and 600 µl of Optimem 1 was placed below each insert. Transfection was performed according to the methods described above.

2.3.7 Harvesting cells for FACS

For 24-well plates:

Cells were washed once with 1 ml PBS per well. 1 ml of a 1:1 Trypsin: Versine mixture was added to each well. The plate was then placed in a humidified 37°C, 5% CO_2 incubator for approximately 10 minutes, until cells had detached from the plastic and dissociated (confirmed by visual examination under the microscope). It was often necessary to pipette the cells through a fine tipped pastette to achieve a single-cell suspension. The suspension was transferred to a 15-ml Falcon tube.

1 ml of Fetal Calf Serum was added to each tube of cells, to inactivate the Trypsin. The cells were spun down in a Jouan centrifuge for 5 minutes at 1,000 RPM.

The supernatant was removed, taking care not to disturb the pellet of cells. 1 ml of PBS was added to each 15-ml Falcon tube to wash the cells and the tube was flicked several times to roughly re-suspend the pellet. The cells were centrifuged again as above. The cells were then re-suspended in 0.5 – 1 ml of PBS, and transferred to a sterile 5-ml polystyrene tube.

For air interface primary tracheal cultures:

600 µl of PBS was placed below the insert and 200 µl of PBS was placed above the insert. Electrical resistance readings were taken before harvesting cells for FACS (see above).

600 µl of a 1:1 Trypsin: Versine mixture was placed below the insert and 200 µl was placed above the insert. The cells were then incubated, harvested, and processed as described above for 24-well plates (see above).

2.3.8 FACS

A Becton Dickinson FACScan machine was used to analyse the cells. The following settings were used:

FSC:
Data Mode: Linear
Detector level: E-1
Amplifier level: 5.8

SSC:
Data Mode: Linear
Detector level: 297
Amplifier level: 1.00

FLH-1, FLH-2 and FLH-3:
Data Mode: Logarithmic
Detector level: 438

10,000 cells were counted for each sample, unless otherwise stated.

2.3.9 βgalactosidase luminometer assay

Cells in a 24-well plate were washed twice with 500 µl of PBS. 150 µl of Lysis buffer (see Recipes) was added to each well, and incubated for one to two minutes at room temperature to allow lysis to proceed. The solution was then transferred to a 96-well plate, and spun in a Jouan CR422 centrifuge, with plate buckets, at 1200 RPM for 5 minutes, to pellet debris. The supernatant (containing βgalactosidase) forms the sample for the subsequent assay (it was sometimes necessary to dilute the sample 1/10 or 1/20 in Lysis buffer to prevent an overload reading in the assay). βgalactosidase enzyme was diluted in Lysis buffer, to make a series of twofold dilutions (from 12,500 ng/ml to 390.63 pg/ml), for use as controls.

200 µl of Reaction buffer (see Recipes) was added to a set of 5-ml polystyrene tubes. 40 µl of sample (or βgalactosidase dilution control) was added to each tube. In addition, negative controls (containing 40 µl lysis buffer, in place of the sample)

were included. The mixtures were shaken and incubated in the dark at room temperature for one hour.

An EG+G Berthold Lumat LB 9507 luminometer was used to assay the βgalactosidase concentration. A program was created to direct the luminometer to inject 300 µl of Accelerator buffer (see Recipes) into the sample, and then read the absorbance of the solution at 530nm. The concentration of βgalactosidase in each sample was calculated by plotting the readings for the controls of known concentration.

2.3.10 βGalactosidase ELISA

The β-Gal ELISA colorimetric enzyme immunoassay kit (Boehringer-Mannheim, cat # 1 539 426) was used according to the manufacturers instructions.

2.3.11 Obtaining RNazol lysates for RT-PCR

Cells in a 24-well plate were washed once with 1 ml of PBS, and then 1ml of RNazol (Ambion, cat #C5104) was added to each well. The plate was incubated for 5 minutes, then the RNazol was pipetted up and down to ensure complete lysis of the cells and collection of the total DNA content. The samples were transferred to 1.5-ml screw-top microcentrifuge tubes and stored in a –70 °C freezer.

2.3.12 Preparing samples for electron microscopy

The Costar inserts were washed 5X in PBS. 200 µl of Fix (2.5% gluteraldehyde in PBS) was applied above the membrane, and 600 µl below. The cells were fixed for at least two hours. The inserts were washed with PBS three times, and then the membrane (with the cell monolayer still attached) was carefully removed from the plastic insert support with a scalpel. The remainder of the processing was done as a service by Steve Mitchell at the Department of Veterinary Medicine, University of Edinburgh.

2.3.13 Statistics

In order to determine the significance of any differences between treatments, statistical tests were employed.

2.3.13.1 T-Test

For phosphoimager analysis of RT-PCR results, βgalactosidase assays, and FACS analysis, a two-tailed T-test was employed. This statistic tests the null hypothesis: $\mu 1=\mu 2$, in other words, the probability that there is no significant difference in the observed values for two treatments, by measuring signal/noise (the T-test is appropriate due to small sample size). The T-value is calculated according to the formula:

$$T = (\mu_1 - \mu_2)/SE(\mu_1 - \mu_2)$$

$$SE(\mu_1 - \mu_2) = \sqrt{([\sigma^2_1/\{n_1-1\}]+[\sigma^2_2/\{n_2-1\}])}$$

Where
μ = mean
σ^2 = variance
n = sample number
degrees freedom = n-1

A statistical table (SurfStat statistical tables website) was then used to calculate the probability (P-value) that the null hypothesis is correct from this T-value. If the P-value was less than 0.05 (e.g. less than 5% probability that the means are the same), then the difference was considered to be significant, and the null hypothesis was rejected. Only significant differences are reported in the text (along with their corresponding P-values). The T-values for these calculations can be found in Appendix F.

2.3.13.2 Test statistic $\mu 1\mu 4 = \mu 2\mu 3$

A more complex statistical analysis was performed by Dr. Naomi Wray (Statistician, MMC, University of Edinburgh), in order to determine whether the observed difference in the average fluorescence intensity following transfection with the $_{PAC}$65kbcfproEGFP vector and the p1kbcfproEGFP vector was proportional to the difference between the $_{PAC}$RC2cmvEGFP vector and the pEGFP-N vector.

We want to test the null hypothesis: $\mu 1/\mu 2 = \mu 3\mu 4$, where:

$\mu 1$ = the average fluorescence intensity of pEGFP-N
$\mu 2$ = the average fluorescence intensity of $_{PAC}$RC2cmvEGFP
$\mu 3$ = the average fluorescence intensity of p1kbcfproEGFP
$\mu 4$ = the average fluorescence intensity of $_{PAC}$65kbcfproEGFP

However, this hypothesis is non-trivial. The null hypothesis: $\mu 1\mu 4 = \mu 2\mu 3$ is equivalent and is easier to test. The triplicate (or duplicate) samples are treated as random effects, as they are sampled from separate populations (e.g. wells) and we therefore expect real differences in the sample means.

The test statistic is calculated according to the formula:

$$T = \frac{X_1 X_4 - X_2 X_3}{\sqrt{([1/n_1 n_4][X_4^2\sigma_1^2 + X_1^2\sigma_4^2 + \sigma_1^2\sigma_4^2] + [1/n_2 n_3][X_2^2\sigma_3^2 + X_3^2\sigma_2^2 + \sigma_2^2\sigma_3^2]}}$$

where:

n = number of sample replicates
X = the mean of the average fluorescence intensity
σ^2 = the Mean Square of the replicates

for the vectors:

pEGFP-N (1)
$_{PAC}$RC2cmvEGFP (2)
p1kbcfproEGFP (3)
$_{PAC}$65kbcfproEGFP (4)

The results of this analysis are discussed in Chapter 4, and the original T-values are reported in Appendix F.

2.4 Recipes

2.4.1 General recipes

Ficoll loading dye
1g Ficoll
100 µl 0.5 M EDTA
100mg Orange G
H_2O → 10ml

Denature solution
1000g 5M NaOH
1461g 5M NaCl
H_2O →5L

Neutraliser Solution
302.75g Tris Base
876.6g NaCl
H_2O → 5L
(adjust pH to 4.5/5 with concentrated HCl)

50 X TAE
1210g Tris Base
285.5ml Glacial Acetic Acid
500ml 0.5M EDTA
H_2O → 5L

20 X TBE
1080g Tris Base
550g Boric Acid
400ml 0.5M EDTA (pH 8.0)
H_2O → 5L

TE
5ml 1M Tris (pH 8.0)
100µl 0.5 M EDTA
H_2O → 500ml
(adjust pH to 8.0 with a few drops of NaOH)

SSC
876.5g NaCl
441.0g Sodium Citrate
H_2O → 5L
(adjust pH to 7.0 with a few drops of NaOH)

10 X TNE
10ml 1M Tris (pH 8.0)
2ml 0.5M EDTA
12g NaCl
H2O →100ml

Hybridisation Mix
125ml 20X SSC
0.5g pyrophosphate (PPi)
12.5ml 20% SDS
50ml Denhart's solution
H_2O → 500ml

50X Denhart's Solution
5g ficoll
5g PVP
5g BSA
H_2O →500ml

PCI
25ml Phenol
24ml Chloroform
1ml Isoamyl Alcohol

Buffer M (Roche)
100 mM Tris-HCl
100 mM KAc
500 mM NaCl
10 mM Dithioerythritol (DTE)

2.4.2 Bacterial recipes

L-Broth
50g Tryptone
25g Yeast Extract
25g Sodium Extract
H_2O →5L
(adjust pH to 7.2 with NaOH)

L-Agar
50g Tryptone
25g Yeat Extract
50g NaCl
H_20 → 5L
(adjust pH to 7.2 with NaOH)
aliquot 400ml + 6g Bacto Agar to each bottle

GTE (minipreps)
0.5g glucose
1.25ml 1M Tris (pH 8.0)
5ml 0.5M EDTA
H_2O → 50ml

+2mg/ml lysozyme
+10ml/ml RNase
(added immediately before use)

GYT
SOB-Mg
10g Bacto-Tryptone
2.5g Bacto yeast extract
5ml 1M NaCl
1.25 ml 1M KCl
H_2O → 500ml

2.4.3 Tissue Culture Recipes

COS7 Growth Media
DMEM (GIBCO, cat #41965-039)
10% Fetal Bovine Serum (GIBCO, cat#10106-151)
1% Pennicillin/ Streptomycin (Stock 10,000u/ml,10,000µg/ml GIBCO, cat#15140-122)

MDCK-IOWA Growth Media
1:1 DMEM: HAM's F12 (GIBCO, cat#21765-029)
10% Fetal Bovine Serum
1% Pennicillin/ Streptomycin

T84 Growth Media
1:1 Vitacell Modified DMEM (ATCC, cat # 30-2002): Ham's F12
10% Fetal Bovine Serum
1% Pennicillin/ Streptomycin

HBE Growth Media
MEM (GIBCO, cat# 31095-029)
10% Fetal Bovine Serum
1% Pennicillin/ Streptomycin

CaCO2 Growth Media
MEM
2mM Glutamine (200mM stock, GIBCO, Cat #25030-024)
1 X non-essential amino acids (100X stock, GIBCO, cat # 11140-035)
1% Pennicillin/ Streptomycin

HBS Growth Media
1ml 1M HEPES
1.5ml 5M NaCl
UHP dH_2O → 50ml
(adjust pH to 7.4, filter sterilise in sterile safety cabinet)

Ca^{2+} Mg^{2+} free MEM
1L UHP dH_2O
3.7g $NaHCO_3$
4g KCl
6.4g NaCl
125mg $NaH_2PO_4.H_2O$
100µg $Fe(NO_3)_{3\text{-}9}$
110µg Sodium Pyrivate (SIGMA, cat#58636)
15mg Phenol Red
12ml Penn/Strep (Stock 5,000u/ml, 5,000µg/ml)

pH to 7/5 with a few drops of concentrated hydrochloric acid. Store at 4°C and filter through a 0.2 µm filter before use.

Dissociation Media
60 ml Ca^{2+} Mg^{2+} free MEM
84 mg Pronase (Boeringer-Mannheim 165121)
6 mg DNase (~3,000u)(Sigma DN-25)

Filter through a 0.2µm filter before use, then add 0.25 µg/ml Amphotericin B and 100 µg/ml Vancomycin.

Airway Media
1:1 DMEM: (GIBCO 41966-029) to HAM's F12 (GIBCO 21765-029)
5% FCS
2% Penicillin/ Streptomycin Stock (Stock 5,000 u/ml, 5,000 µg/ml)
600 µl Human Actrapid Insulin (Stock 100iu/ml)

Store in flasks with vented caps in a 37 °C, 6% CO_2 tissue culture incubator, to allow the media to breathe for CO_2 exposure. Before use add 0.25 µg/ml Amphotericin B.

USG Media
233 ml DMEM(GIBCO 41966-029): 247 ml HAM's F12 (GIBCO 21765-029)
10 ml Lyophilised Ultroser G serum substitute (GIBCO 15950-017)
10 ml Penn/Strep (Stock 5,000u/ml, 5,000µg/ml)

Store at 4°C. Before use, add 0.25 µg/ml Amphotericin B.

Bgal luminometer assay Lysis buffer
9.15 ml 100 mM K_2HPO_4
0.85 ml 100 mM KH_2PO_4
20 µl Triton X (SIGMA, cat#T-9284)
10 µl DTT

Bgal luminometer assay Reaction Buffer
9.5 ml Na_2HPO_4
0.5 ml NaH_2PO_4
10 µl 1M $MgCl_2$
100 µl Tropix Galacton (Tropix, cat #GC020)

Bgal luminometer assay Accelerator
1 ml Tropix Emerald (Tropix, cat#LAE250)
2 ml 1M NaOH (stock must be made up fresh)
7 ml H_2O

Chapter 3: Vector Design and Construction

3.1 Introduction:

The first generation of vectors designed for CF gene therapy contained the CFTR reporter gene, coupled to a strong promoter (e.g. the P_{CMV} viral promoter/enhancer). However, a more elegant design may be ultimately required for gene therapy vectors, incorporating regulatory elements to restrict expression to the correct levels and to the correct cell types. Our approach is to utilise the endogenous CFTR control elements in genomic context vectors (GCVs), to regulate expression (see Section 1.2).

In order to test the effects of such regulatory elements in a gene therapy context, a series of vectors was created, coupling different portions of the CFTR genomic region to the EGFP reporter gene. Regions 5' of the CFTR gene were included in some of these vectors, to test the effects of proximal and distal upstream regulatory elements upon expression. CFTR intron 1 was included in one vector, to test the effects of the intron 1 DHS upon expression, as previous reports have demonstrated that this element upregulates CFTR expression in cells of the gut epithelium (Smith, AN et al, 1996; Rowntree et al, 2001). These regions were linked to the EGFP reporter gene either directly, or by an internal ribosome entry site (IRES). An IRES allows translation to initiate from an internal site, rather than the 5' end of the mRNA. A previous report (Vassaux and Huxley, 1997) has demonstrated that CFTR transcriptional activity can be accurately monitored by inserting an IRES and reporter gene downstream of the CFTR coding sequence.

Two small plasmids, p1kbcfproEGFP and pCFTRiresEGFP, were generated by conventional restriction/ligation reactions. However, it is not easy to manipulate large PAC constructs by such conventional cloning procedures, and thus a recombinogenic method was used to create the PAC constructs $_{PAC}$65kbcfproEGFP, $_{PAC}$RC2cmvEGFP, $_{PAC}$RC1iresEGFP, and $_{PAC}$RC2iresEGFP. The foldout Appendix E shows a diagram of the expression cassette in each of these vectors (and some control constructs), for easy reference. The predicted sequence of each vector was assembled using the HGMP GCG 'assemble' and 'seqed' software.

Recombinant clones were detected by plasmid restriction, colony PCR, or Southern Blot. These plasmids were then characterised by restriction analysis, to confirm their structure. When a PCR step was included in the construction process, the appropriate region was sequenced, to ensure that a deleterious mutation had not been introduced.

In sum, a panel of vectors was created, to provide the tools for subsequent transfection studies. Upon transfection into mammalian cells, these vectors will report upon the effects of CFTR 5' and intron 1 elements on EGFP transgene expression.

3.2 Recombinogenic method (Fig. 3.1):

A PCR product was generated with a loxP site at the 3' end and a region homologous to the site of vector insertion at the 5' end. These features can theoretically be introduced by primer-directed mutagenesis if they are not already present in the region to be amplified, making the system highly flexible[1].

Two recombination events were used to fuse the PCR product with the vector (Fig. 3.1):

1. **Cre/loxP recombination.** The Cre enzyme was used to catalyse a recombination event between loxP sites in the parent vector and PCR product *in vitro*. This generates a linear molecule.

2. **Homologous recombination.** This was accomplished by transforming the linear molecule into DH10B-u/pSpRecGam cells. The DH10B-u/pSpRecGam strain exhibits a high frequency of homologous recombination, and was made from the *E. coli* strain DH10B-u by introducing a plasmid (pSpRecGam)

---

[1] This method is sound in principle, however it may not prove possible to incorporate the loxP site by primer-directed mutagenesis, as this sequence is palindromic and is predicted to form a strong secondary structure in the primer. The practicality of this approach must be tested.

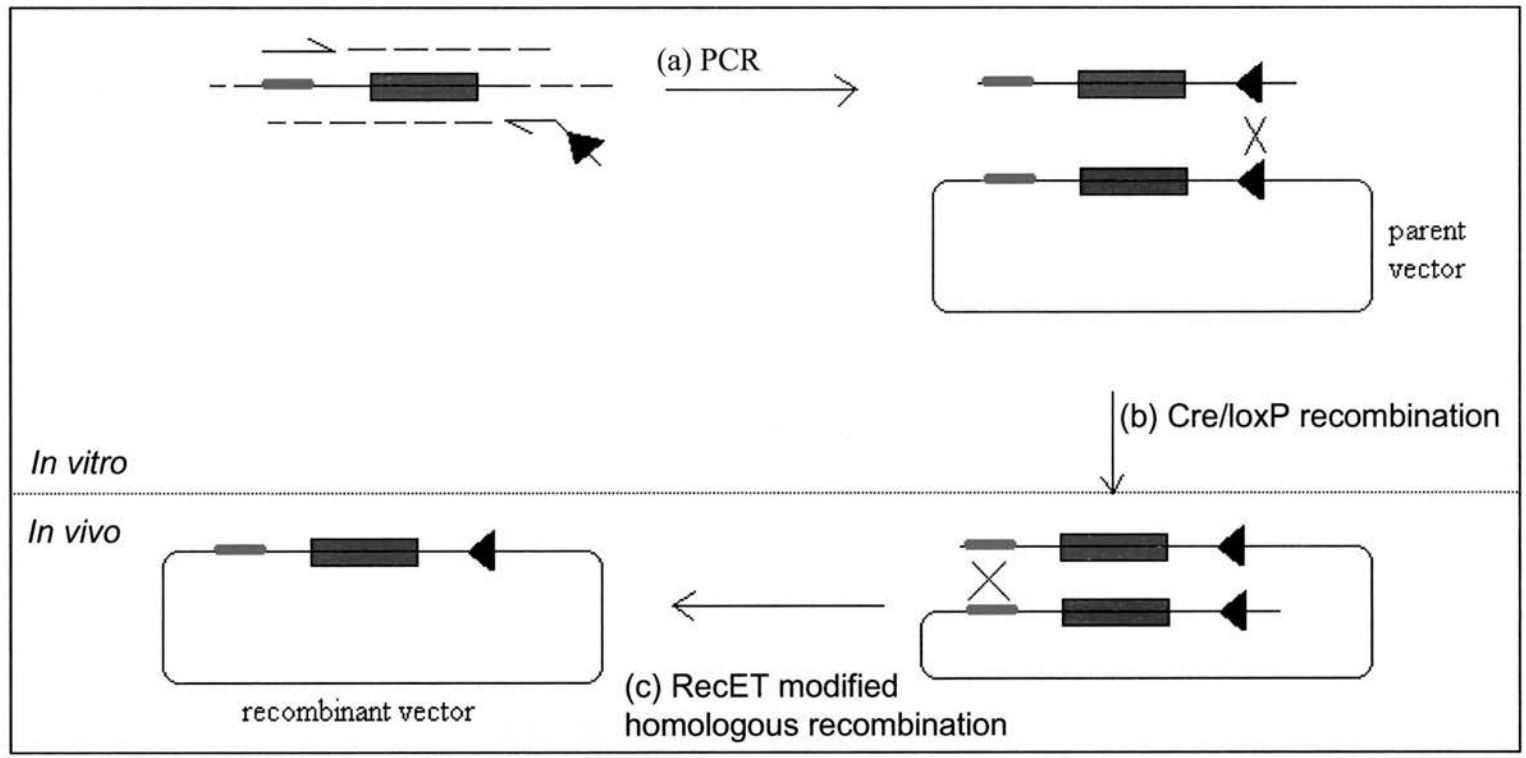


Figure 3.1. Recombinogenic method of vector construction. (a) PCR generates a fragment containing a loxP site (◄) and a region of homology to the parent vector (▬). **(b)** Cre recombinase catalyses a recombination event between the lox P sites in the fragment and the parent vector. **(c)** The linear vector is transformed into the pSpRecGam/DH10B-u E. coli strain, where a homologous recombination event occurs to create the recombinant vector. ▬ represents the region that will be introduced to the vector by this double recombination method, ▬ represents the region that will be removed

encoding RecET and Gam. The RecET protein complex promotes RecA-independent homologous recombination. Gam increases the efficiency of recombination by repressing recBCD, which digests linear DNA. Replication of the pSpRecGam plasmid is dependent on the presence of IPTG (which represses $LacI^q$ and thus derepresses pSpRecGam replication), thus the high frequency of homologous recombination can be turned off after the plasmid is introduced by exclusion of IPTG to prevent further aberrant recombination events. The use of RecET for plasmid construction was reported by Zhang (Zhang et al, 1998a). This event circularises the recombinant vector.

This double-recombination method was reported by Boyd and colleagues (Boyd et al, 1999b). The use of a PCR product in lieu of a restriction fragment for insertion, and the use of primer-directed mutagenesis to incorporate a loxP site for recombination, are refinements made for this work.

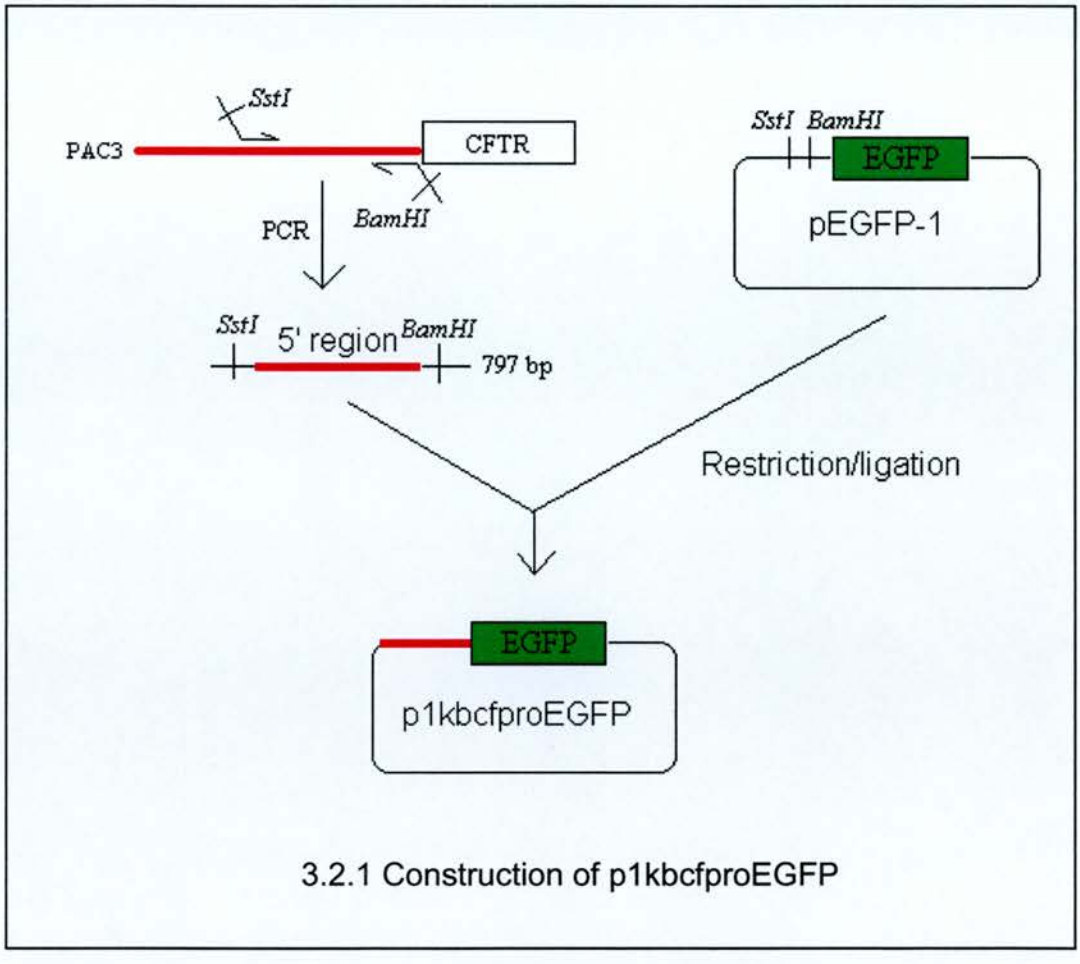
3.3 Results

3.3.1 p1kbcfproEGFP

The small plasmid p1kbcfproEGFP contains 797 bp of CFTR 5' region, upstream of the EGFP reporter gene (see fold out Appendix E). This vector will report upon the effect of proximal CFTR 5' promoter elements on transgene expression. A conventional cloning procedure was used to make this plasmid (Fig. 3.2.1).

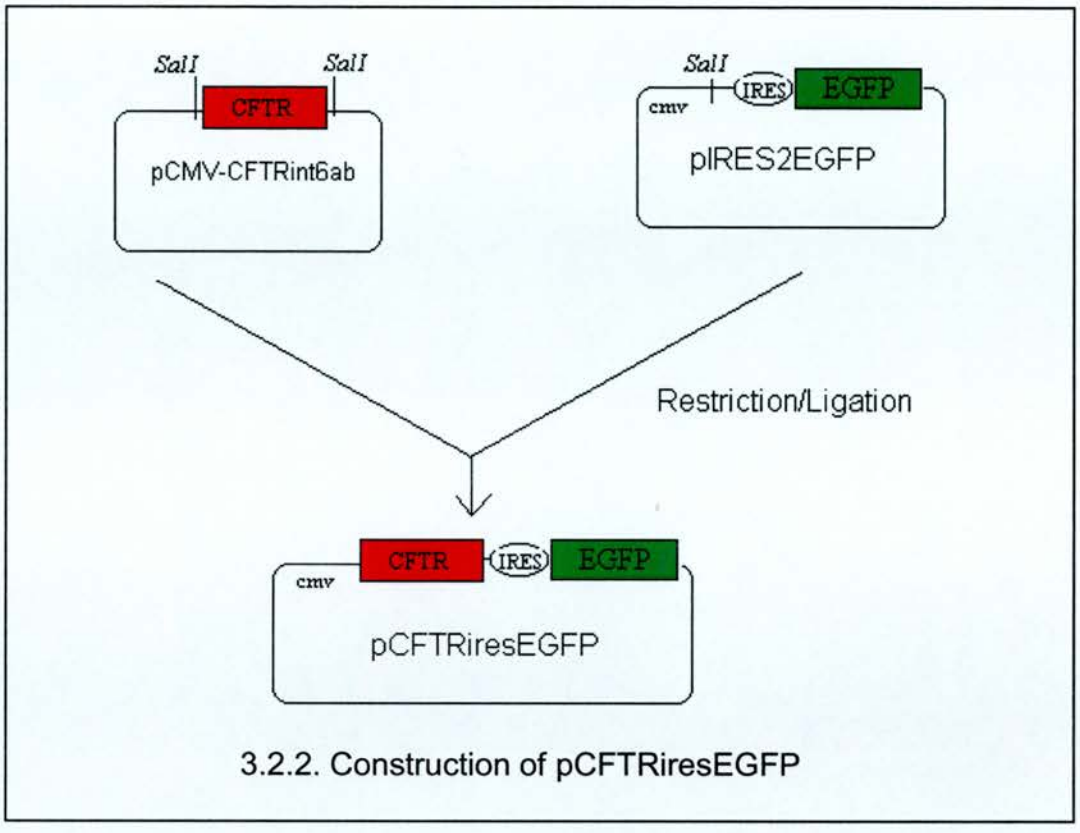
797 bp of sequence 5' of the ATG start site of the CFTR gene was amplified from $_{PAC}$3 (Boyd and Porteous, 1997), by PCR. The primers used in this reaction, 5'cfpro and 3'cfpro, were designed to introduce restriction sites *BamHI* and *SstII*, respectively, onto the ends of the PCR product, allowing subsequent restriction/ligation reaction (see Appendix A for primer sequences). A hot-start program was used to enhance amplification from the PAC template (Chou et al, 1992; see Methods). The Expand™ High Fidelity PCR system (Roche) was used in this reaction; this enzyme blend has proofreading (3'-5' exonuclease) activity, minimising the possibility of PCR mutagenesis (Roche website). The PCR product

Figure 3.2. Vector construction. This figure shows the schematics of vector construction for **(3.2.1)** p1kbcfproEGFP, **(3.2.2)** pCFTRiresEGFP, **(3.2.3)** $_{PAC}$65kbcfproEGFP, **(3.2.4)** $_{PAC}$RC2cmvEGFP, **(3.2.5)** $_{PAC}$RC1iresEGFP, and **(3.2.6)** $_{PAC}$RC2iresEGFP. ▬ Represents CFTR 5' region, ▬ represents EGFP, ▬ represents CFTR, ▬ represents βgal. (ires) represents an internal ribosome entry site.

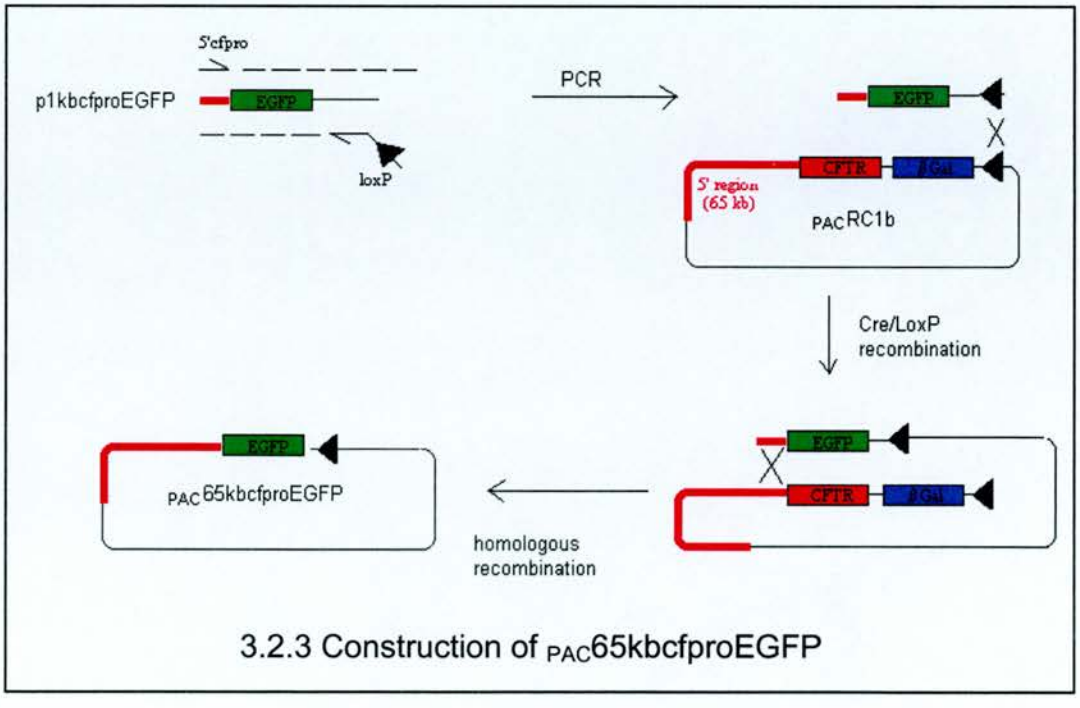


3.2.1 Construction of p1kbcfproEGFP

Figure 3.2 Continued.

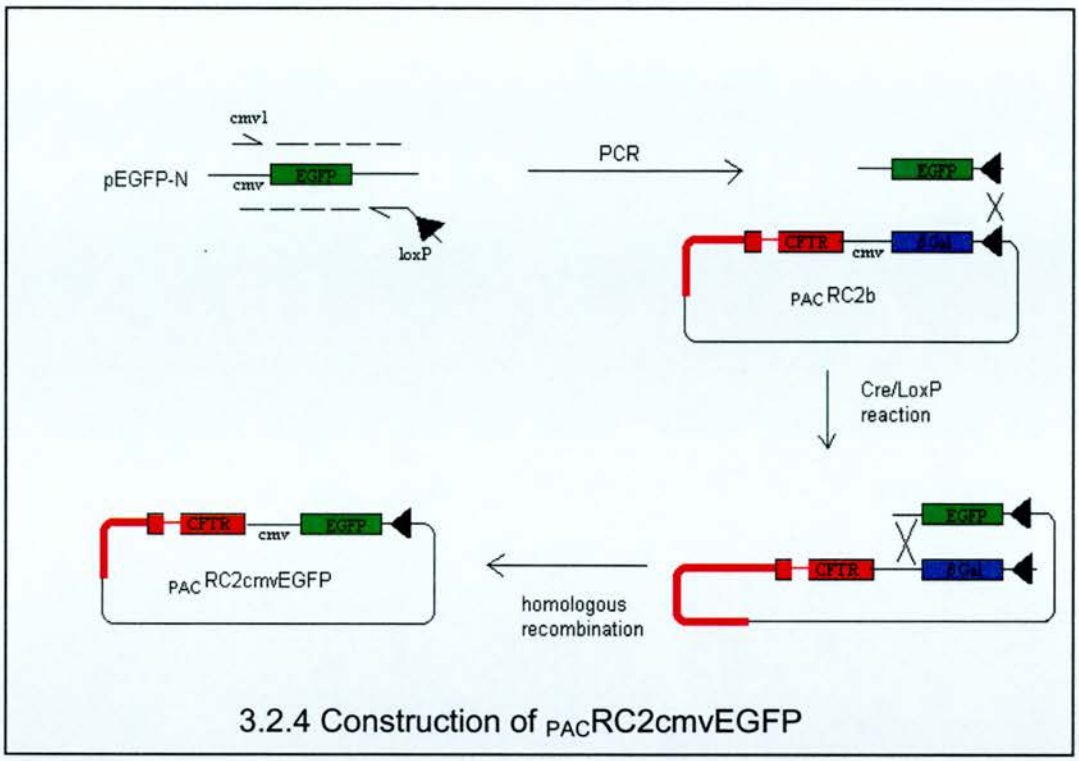


3.2.2. Construction of pCFTRiresEGFP

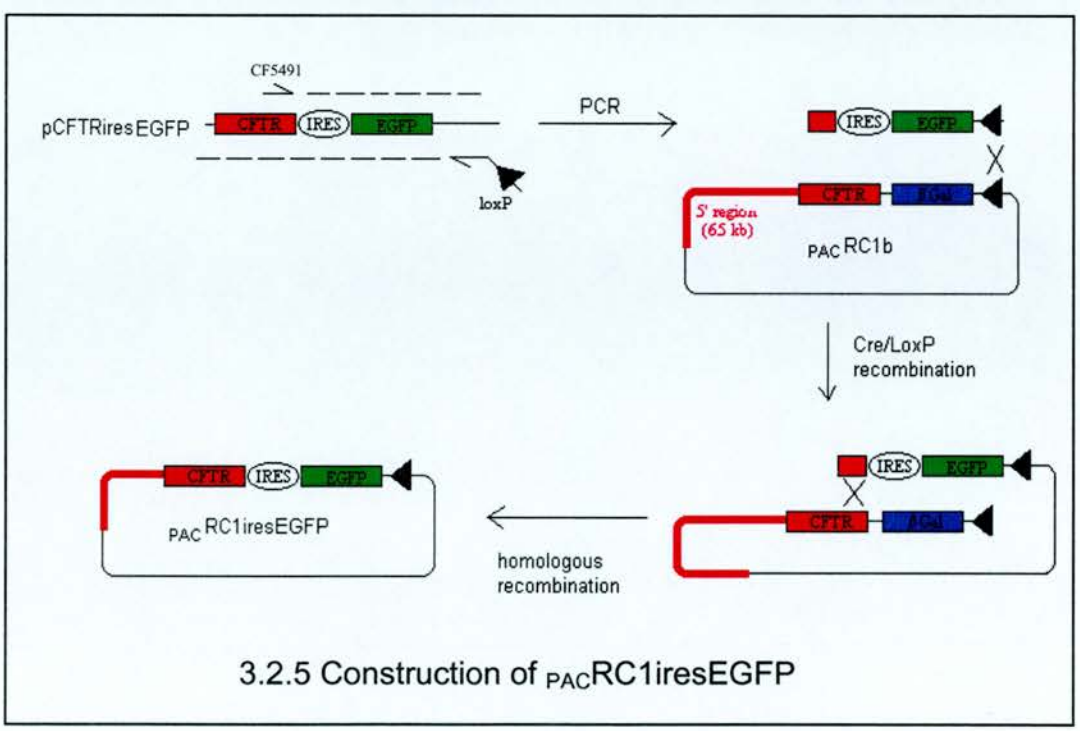


3.2.3 Construction of $_{PAC}$65kbcfproEGFP

Figure 3.2 Continued.

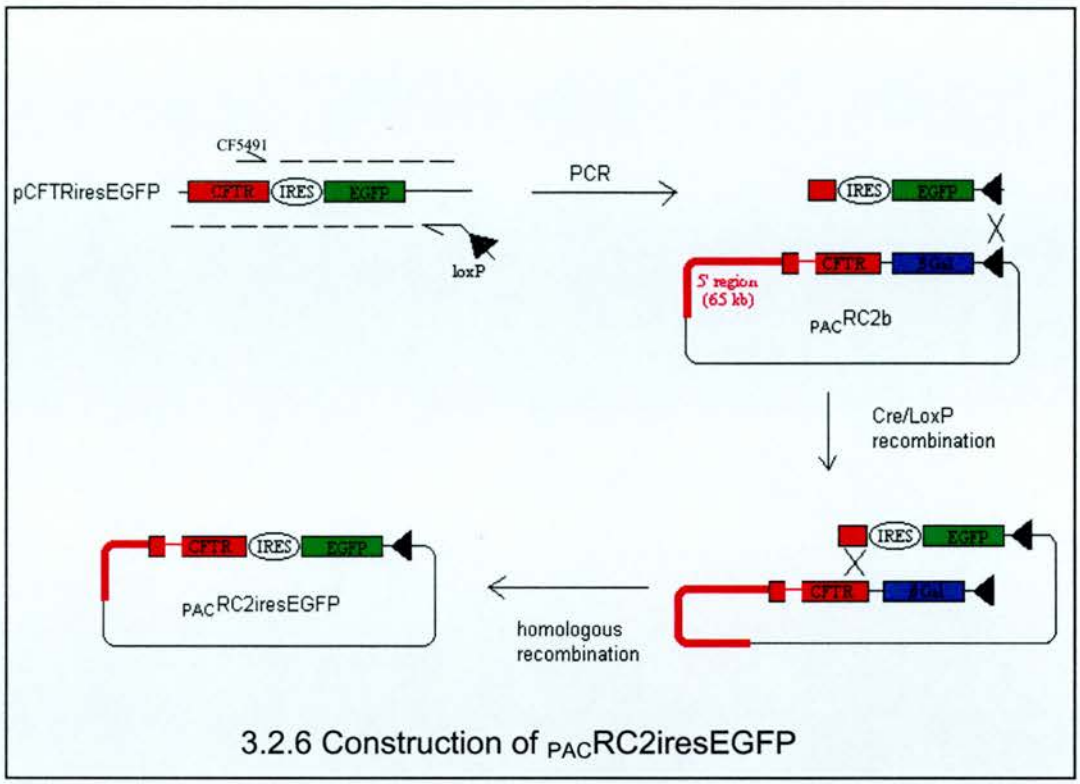


3.2.4 Construction of $_{PAC}$RC2cmvEGFP



3.2.5 Construction of $_{PAC}$RC1iresEGFP

Figure 3.2 Continued.



was digested with *SstII* and *BamHI*, and purified by ethanol precipitation (see Methods).

pEGFP-1 is a promoterless vector, designed to monitor transcription from different promoter/enhancer combinations inserted into the multiple cloning site (MCS) located upstream of the EGFP coding sequence (see Appendix D, and Clontech website). This vector was digested with *SstII* and *BamHI* (proximal restriction sites in the multiple cloning site), and treated with phosphatase to prevent re-ligation of the parent vector (see Methods).

SstII produces a 3' overhang (GC), while *BamHI* produces a 5' overhang (CTAG). Thus, the vector and insert will have compatible 'sticky ends', allowing the two to be joined by ligation (see Methods). A series of ligation reactions were set up with different ratios of plasmid DNA to PCR product (Table 3.1 – p.124).

A 2 µl aliquot of each reaction was disk dialysed and transformed by electroporation into a 20 µl aliquot of DH10B-u *E. coli* competent cells (see Methods). The time constants were between 1.9 and 2.0 ms. Cells were plated out on L-agar plates containing Kanamycin and grown overnight (see Methods). The following day the plates were scored for colony growth (Table 3.3a – p.125).

40 colonies were chosen at random and patched onto a new plate, and then crude DNA minipreps were performed (see Methods- two of these, samples #19 and 30, did not grow, and were excluded from further analysis). *DraI* and *BssHII* digests were performed to identify recombinant clones.

Six colonies (# 3,23,25,33,34 and 37) showed the correct digest pattern for the recombinant plasmid with both enzymes (Fig. 3.3). 24 of the remaining colonies showed the correct digest pattern for the parent vector. Three colonies (# 26,28 and 29) showed the correct digest pattern for the recombinant plasmid, except for an extra faint band in the *BssHII* digest. Although this band could have been an artifact, these clones were excluded from further analysis. Five colonies (# 2,6,21,36 and 38) presented unusual digest patterns with one, or both enzymes, no further analysis was performed on these errant clones.

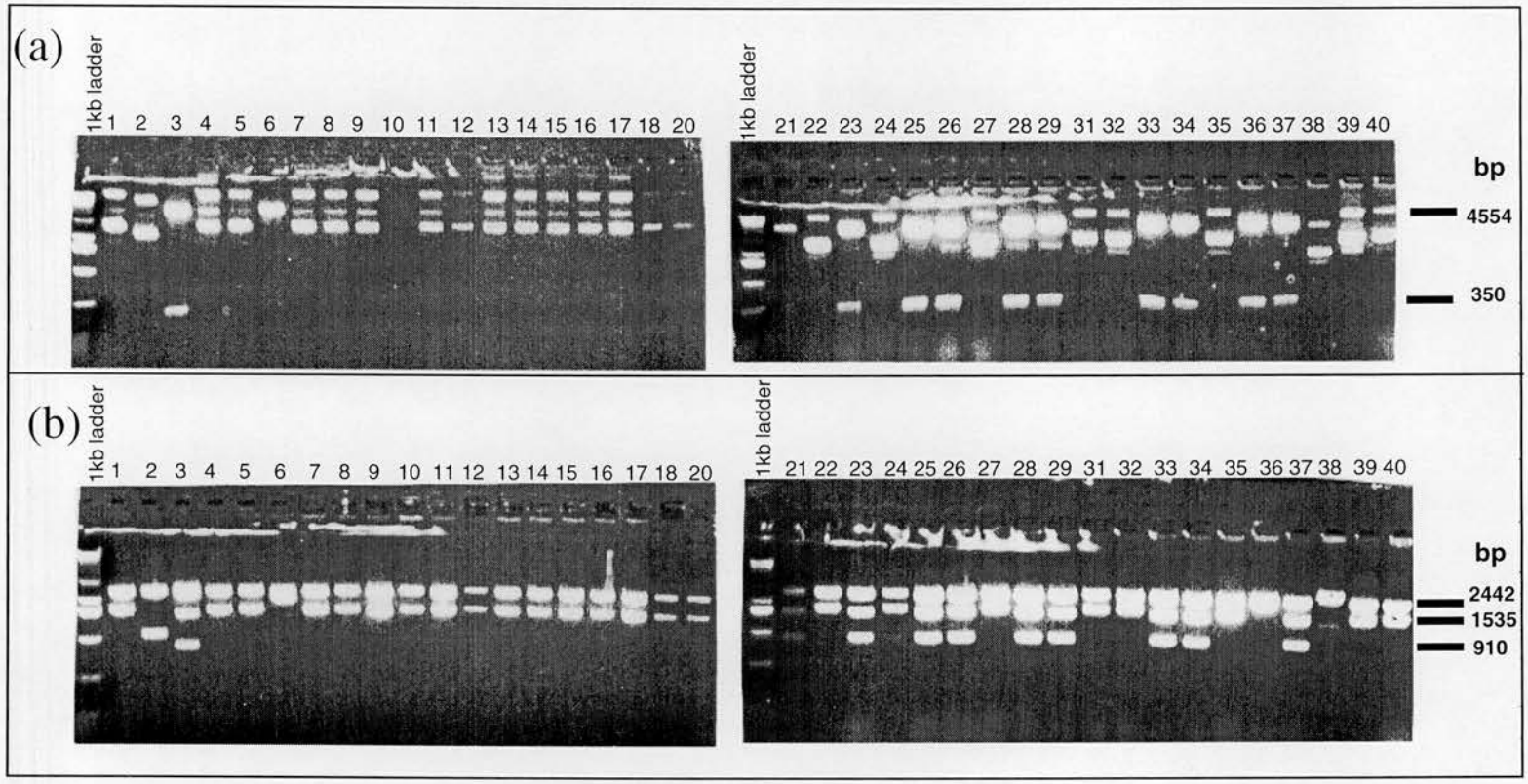


Figure 3.3. Diagnostic digests of p1kbcfproEGFP colonies. Miniprep DNA from p1kbcfproEGFP colonies(1-40, as labelled) was digested and run down an agarose gel (see Methods). **(a)** *BssHII* digestion produced the correct pattern (4554 bp and 350 bp bands) in colonies 3, 23, 25, 33, 34, 36, and 37. **(b)** *DraI* digestion produced the correct pattern (2442 bp, 1535 bp, and 910 bp bands) in colonies 3, 23, 25, 26, 28, 29, 33, 34, and 37.

Qiagen maxiprep DNA was prepared from the six clones #3, 23, 25, 33, 34, and 37. This DNA was used as a template to sequence across the insertion site, in order to check for mutations derived from the PCR primers, or introduced in the PCR reaction. The vector DNA was sequenced directly using a Rhodamine-sequencing kit (see Methods); the primers seqF, seqR, seqF2, and seqR2 were used in this reaction (see Appendix A for primer sequences). The sequences were run on an ABI-Prism sequencher as a service, by Angie Fawkes in the Clinical Genetics Section, MMC, University of Edinburgh. The Consed software package was used to assemble contigs from the original sequences for each clone, and then the HGMP GCG 'bestfit' software was used to compare these contigs to the predicted vector sequence.

Surprisingly, mutations were identified in several of the clones, despite the use of HPLC-purified primers and the Expand™ High Fidelity PCR system during vector construction. Colony #34 was free from mutations in the inserted region (see Appendix B1 for the sequencing results); this clone was named p1kbcfproEGFP.

3.3.2 pCFTRiresEGFP

The small plasmid pCFTRiresEGFP contains the coding sequence of both CFTR and EGFP. An internal ribosome entry site is located between the genes, allowing them to be translated as separate proteins from a single mRNA (Clontech website). A P_{CMV} promoter drives expression (see fold-out Appendix E). This construct acted a control for transfection studies and was instrumental in the creation of further vectors.

pCFTRiresEGFP was created by a simple restriction-ligation reaction (Fig. 3.2.2). The region containing the CFTR coding sequence was excised from the pCMV-CFTRint6ab vector (Boyd et al, 1999a) by restriction with *SalI*, then gel purified.

The Clontech vector pIRES2EGFP (Clontech website) was designed to make transcriptional fusions with EGFP. A multiple cloning site has been incorporated upstream of an internal ribosome entry site (IRES) for easy insertion of a transgene coding sequence (Clontech website); Appendix D shows a diagram of this vector. The EGFP gene is located downstream of the IRES. This vector was linearised by restriction with *SalI*, which cuts in the multiple cloning site. Linear vector was gel

purified and treated with phosphatase to prevent re-ligation of the parent vector (see Methods).

~125 ng linearised pIRES2EGFP vector and ~ 250 ng of the *SalI* fragment from pCMV-CFTRint6ab were ligated in 10 µl volume (see Methods). A 2 µl aliquot was disk dialysed and electroporated into 30 µl of DH10B-u competent *E. coli* cells. The time constant was 2.2 ms. Cells were plated out on L-agar plates containing Kanamycin and grown overnight (see Methods).

The plates were scored for growth the next day. Every plate had approximately 1,800 colonies. 1,000 of these colonies were chosen at random and patched onto a new plate in grid formation. A colony lift was performed (see Methods) and the filters were probed with a mixture of ^{32}P radioactively-labelled PCR products Ex4-9 and Ex19-24 (see Appendix A, and Methods). Three colonies (#349, 501 and 703) showed a strong, positive signal (Fig. 3.4).

Qiagen DNA maxipreps were prepared from these colonies. Diagnostic restriction digests were performed to identify colonies where the CFTR fragment had inserted in the correct orientation (the insert could have attached in either orientation, due to the identical restriction ends).

One of these (colony #501) showed the correct restriction digest pattern with enzymes *AflII*, *BamHI*, and *XbaI* (Fig. 3.5). This plasmid was named pCFTRiresEGFP.

3.3.3 $_{PAC}$65kbcfproEGFP

The PAC vector $_{PAC}$65kbcfproEGFP contains 65 kb of CFTR 5' region upstream of the EGFP reporter gene. This vector will report upon the effects of distal CFTR 5' elements on expression (see fold-out Appendix E). A double recombination method was used to create $_{PAC}$65kbcfproEGFP (Fig. 3.2.3).

A region containing the 797 bp of CFTR 5' region and the EGFP gene was amplified from p1kbcfproEGFP by PCR (see Methods). The primers 5'cfpro and loxP were used in this reaction (see Appendix A). The loxP primer anneals downstream of the

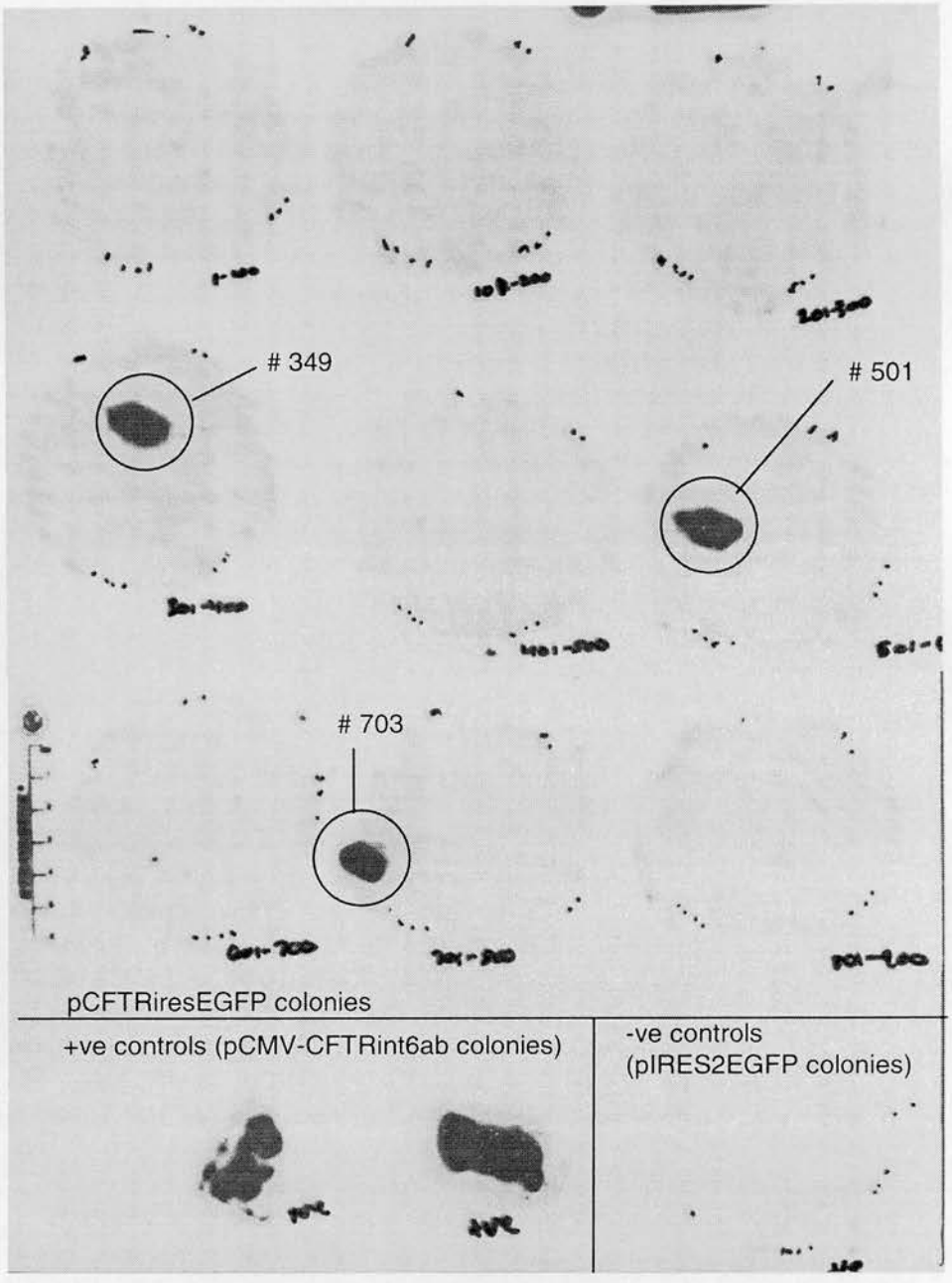


Figure 3.4 pCFTRiresEGFP colony hybridisation. pCFTRiresEGFP clones were patched in a grid formation, along with a positive control (pCMV-CFTRint6ab colonies) and a negative control (pIRES2EGFP colonies - as labelled). A colony lift was performed, and the filters were probed with a mixture of PCR products Ex4-9 and Ex19-24, labelled with ^{32}P-CTP (see Methods). This figure shows an X-ray film, exposed to the filters for 5-hours. A distinct signal can be seen in three colonies (circled): #349, 501 and 703, and in the positive controls. No signal was seen in the negative controls.

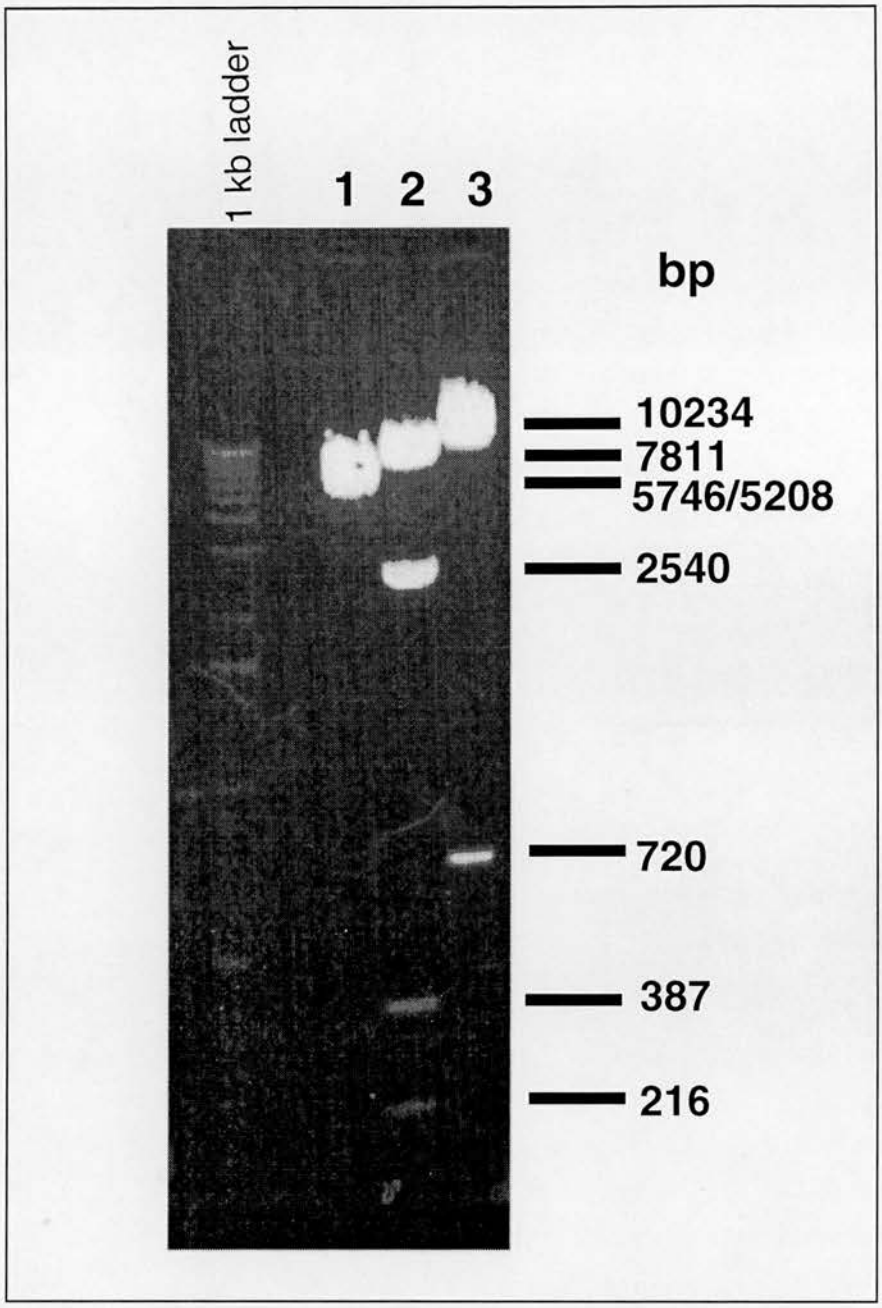


Figure 3.5. Diagnostic digest of pCFTRiresEGFP. Maxiprep DNA from pCFTRiresEGFP colony #501 was digested and run down an agarose gel (see methods). **(1)** *AflII* digestion produced the expected doublet of size 5746/5208 bp **(2)** *BamHI* digestion produced the expected bands of size 7811, 2540, 387 and 216 bp **(3)** *XbaI* digestion produced the expected bands of size 10234 and 720 bp.

EGFP gene and introduces a loxP site onto the end of the PCR product. The Expand™ High fidelity PCR system was used in this reaction (see Methods). The PCR product was gel purified to minimise vector contamination, and purified by ethanol precipitation (see Methods).

An *in vitro* Cre reaction was used to catalyse a recombination event between the loxP sites in the PCR product and $_{PAC}$RC1b vector (this vector is described in Appendix D). Cre reactions (see Methods) were set up with different ratios of vector to PCR product; controls with no PCR product and no Cre were also included (see Table 3.2a – p.124).

2 µl aliquots of these reactions were disk dialysed and transformed by electroporation into 30 µl DH10B-u/pSpRecGam competent *E. coli* cells (see Methods - three aliquots were transformed for reactions 1 – 3, and one aliquot for control reactions 4 + 5). In addition, a 2 µl aliquot of each reaction was transformed into DH10B-u cells as a control. The time constants were between 2.0 and 2.2 ms. Sample 2c was lost.

The cells were plated on L-agar plates containing Kanamycin, IPTG, and Xgal and grown overnight (see Methods). The next day, the plates were incubated at 4 °C for several hours to enhance IPTG/Xgal staining for βgalactosidase. As the recombination event removes the βgalactosidase cassette from the $_{PAC}$RC1b parent vector, recombinant colonies will be white, in contrast to blue $_{PAC}$RC1b colonies.

The plates were scored for growth and colony colour (Table 3.3b – p.126). The Cre reaction plates from DH10B-u/pSpRecGam transformations contained both white and blue colonies, while the DH10B-u plates and Cre controls contained only blue colonies. The 77 white colonies were picked and patched onto a fresh L-Agar plate containing Kanamycin, IPTG, and Xgal (see Methods).

28 of the 77 colonies acquired a blue colour after patching; this could have resulted from misidentification of the original colony colour (small colonies were often a very faint blue colour, easily mistaken for white), or contamination with adjacent blue colonies during the picking process.

Colony PCR was performed on the remaining 49 white colonies, to identify correct clones, using the primers GFPs and GFPas (see Appendix A), which amplify a 442 bp fragment of the EGFP sequence. The products were electrophoresed down a 1% TBE gel (Fig. 3.6). 19 colonies (#3, 13, 18, 24, 30, 36, 39, 41, 48, 49, 50, 53, 54, 56, 59, 61, 62, 63, +75) showed a 442 bp band, suggesting that the EGFP gene was present.

To confirm that these were recombinant colonies (and not, for example, recovered p1kbcfproEGFP template from the PCR reaction) additional colony PCRs were performed. To confirm that the entire insert was present; primer sets 1F/2R and 3F/4R were used (see Appendix A); the resulting PCR products overlap and together span the site of insertion. The products were electrophoresed down a 1% gel (Fig. 3.7). All of the colonies showed the correct size bands for both PCR reactions - a 1038 bp band for 1F/2R and a 1014bp band for 3F/4R (the colony #18 1F/2R sample was lost in this experiment, but this colony produced a band of the correct size in a repeat experiment – results not shown).

Crude DNA minipreps were made from these 19 colonies (see Methods). *ClaI* and *EagI* digests were set up on the miniprep DNA (see Methods); the digests were electrophoresed down a 1% TBE gel. All 19 colonies showed the correct digest pattern for the recombinant clone, in contrast to the $_{PAC}$RC1b parent clone (Fig. 3.8).

Qiagen DNA Maxipreps were performed on three of these colonies - #13,18 and 24. Bulk reactions of the 1F/2R and 3F/4R PCR products were generated using the maxiprep DNA as template. These PCR products were purified with a Qiaquick PCR-purification kit (see Methods). Primers 1F, 1R, 2F, 2R, 3F, 3R, 4F and 4R (see Appendix A) were used to sequence the PCR product with a Rhodamine-sequencing kit (see Methods). It was paramount to sequence the recombinant region to check for mutations introduced by the polymerase chain reaction or small insertions, deletions, or rearrangements resulting from imperfect recombination events.

The sequencing of clones #13 and 18 was stopped when mutations were discovered. Clone #24 was free from mutations in the sequenced region (the sequencing results are included in Appendix B2); this clone was named $_{PAC}$65kbcfproEGFP.

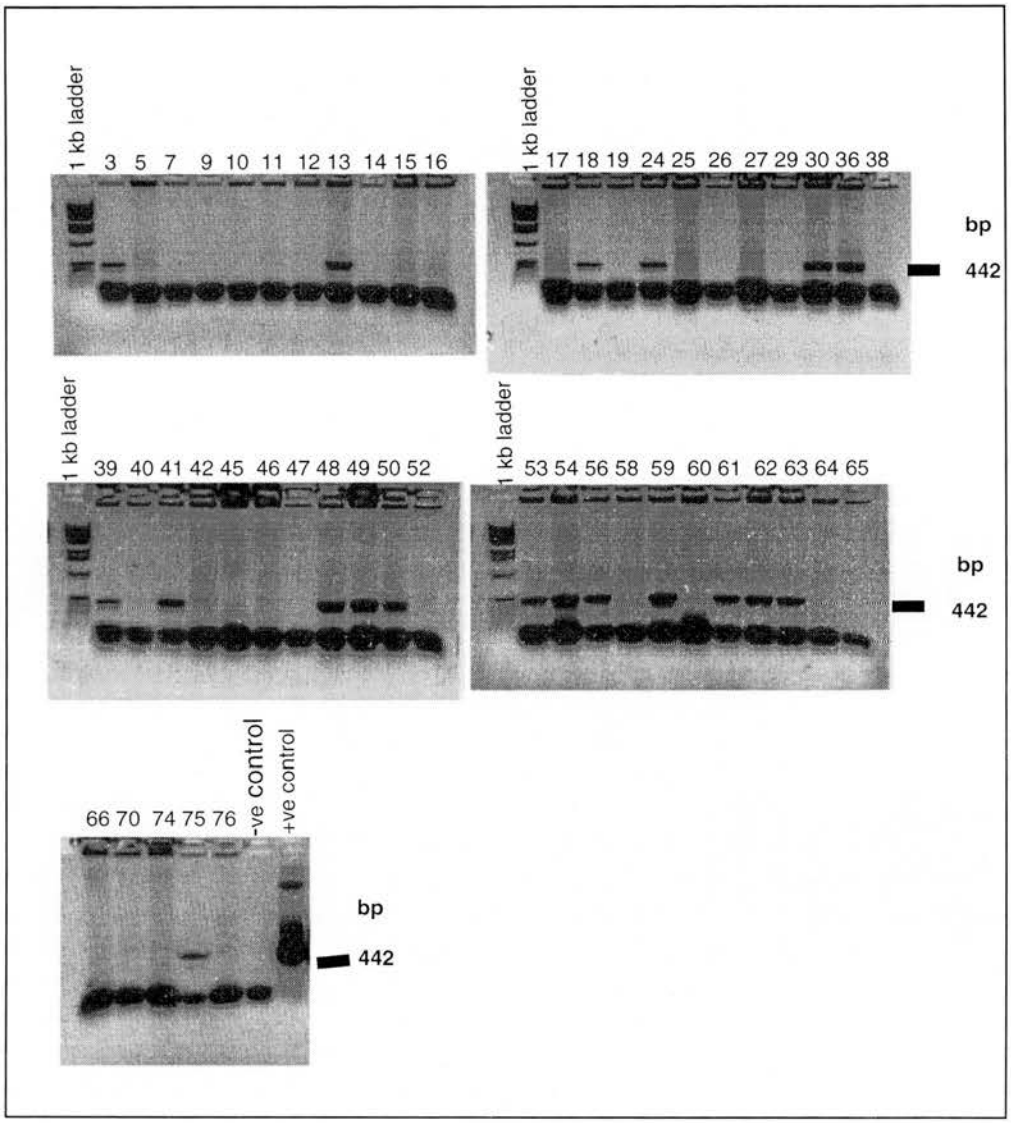


Figure 3.6. GFPs/GFPas PCR on $_{PAC}$65kbcfproEGFP colonies. Colony PCR was performed on $_{PAC}$65kbcfproEGFP colonies 3-76, or a positive or negative control (as labelled), using the primers GFPs and GFPas (Appendix A). Colonies 3, 13, 18, 24, 30, 36, 39, 41, 48, 49, 50, 53, 54, 56, 59, 61, 62, 63 and 75 and the positive control produced a PCR product of size 422 bp, indicating that the EGFP gene is present in these colonies.

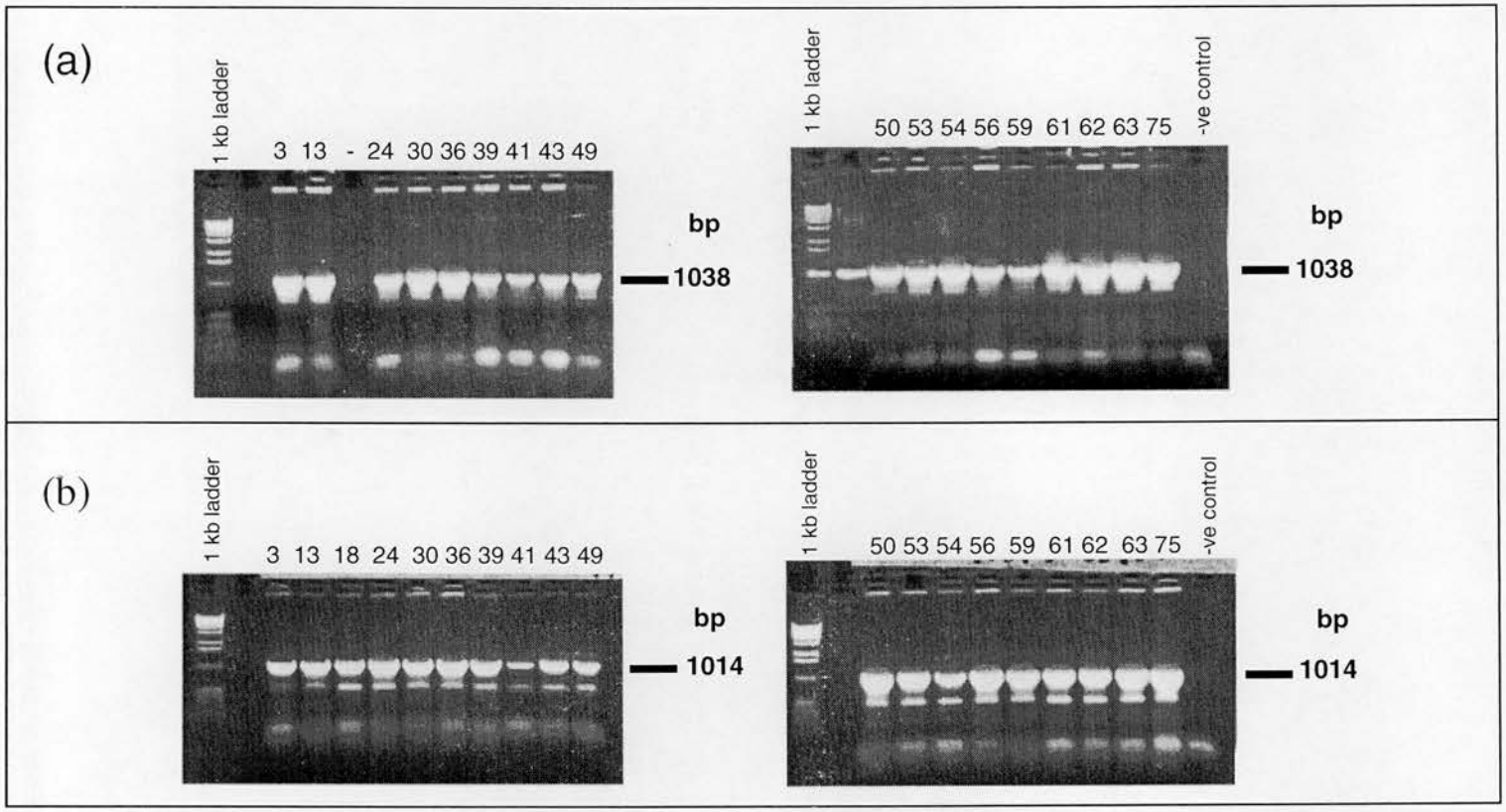


Figure 3.7. 1F/2R and 3F/4R PCR on $_{PAC}$65kbcfproEGFP colonies. Colony PCR was performed on $_{PAC}$65kbcfproEGFP colonies 3-75, or a negative control (as labelled), using the primers 1F and 2R or the primers 3F and 4R (Appendix A). **(a)** For the 1F/2R PCR, every colony produced a 1038 bp band (except colony #18, where the PCR product was lost). **(b)** For the 3F/4R PCR, every colony produced a band of 1014 bp. This indicates that every colony, except possibly #18, contains the entire insert.

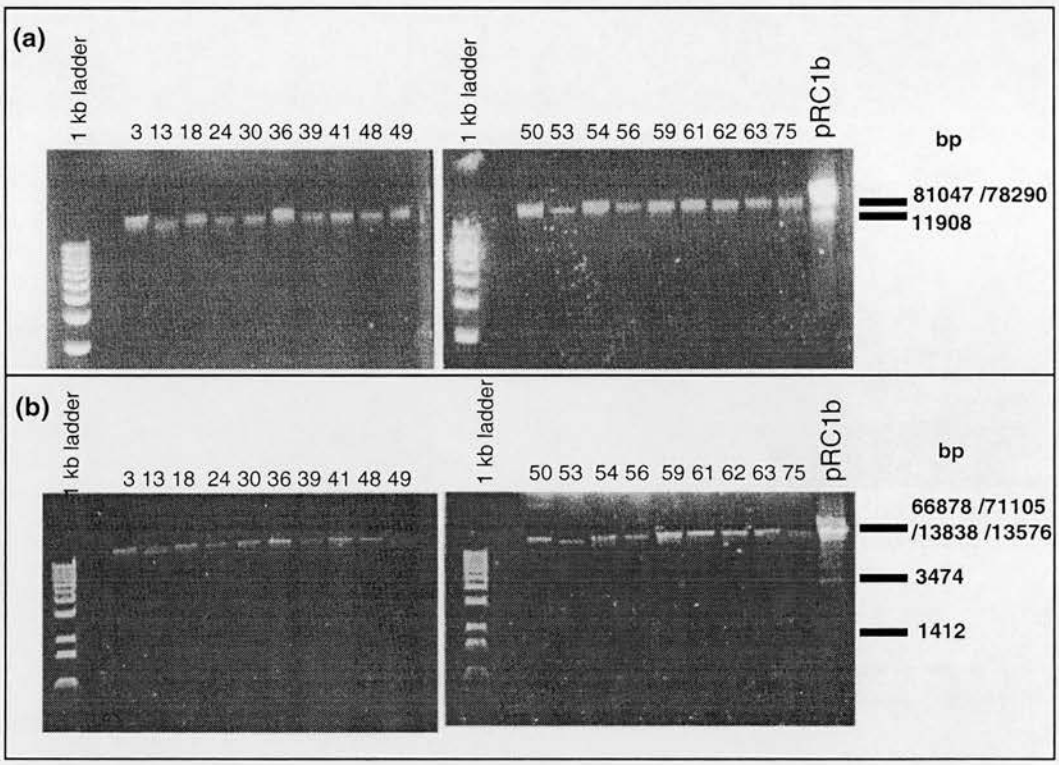


Figure 3.8. Diagnostic digests of $_{PAC}$65kbcfproEGFP colonies. Miniprep DNA from $_{PAC}$65kbcfproEGFP colonies (3-75) and $_{PAC}$RC1b (as labelled) was digested and run down an agarose gel (see Methods). **(a)** *ClaI* digestion of every $_{PAC}$65kbcfproEGFP colony produced a single band of the correct size for the recombinant clone (81047 bp). In contrast digestion of $_{PAC}$RC1b produced two bands (78290 bp and 11908 bp) **(b)** *EagI* digestion of every $_{PAC}$65kbcfproEGFP colony produced a single band (representing a doublet of 66878/13838 bp): the correct pattern for the recombinant colony. In contrast, digestion of $_{PAC}$RC1b produced three bands; a large band (representing the 71105/13576 bp doublet) and two smaller bands (3474 and 1412 bp). The *EagI* digest for both $_{PAC}$65kbcfproEGFP and $_{PAC}$RC1b should also produce a 331 bp band, which is too small to be visualised on this gel.

Finally, restriction digests of $_{PAC}$65kbcfproEGFP were run down a pulse field gel (see Methods). A pulse field gel will separate bands of a larger size, allowing us to confirm that the large PAC vector has not suffered any gross rearrangements. *SacII* and *PmeI* digestion of $_{PAC}$65kbcfproEGFP produced the correct pattern for the recombinant vector, in contrast to the $_{PAC}$RC1b parent vector (Fig 3.9).

3.3.4 $_{PAC}$RC2cmvEGFP

The PAC vector $_{PAC}$RC2cmvEGFP contains the EGFP gene under the control of a P_{CMV} promoter (see fold-out Appendix E). This is an essential control for transfection studies, as it reports on the effects of vector size upon expression from the P_{CMV} promoter.

$_{PAC}$RC2cmvEGFP was produced in a similar manner to $_{PAC}$65kbcfproEGFP (Fig. 3.2.4). A region containing a portion of the P_{CMV} promoter and the EGFP gene was amplified from the vector pEGFP-N (described in Appendix D), using the primers cmv1 and loxP (see Appendix A) and the Expand™ high fidelity PCR system. The PCR product was purified as described above. The pEGFP-N vector was linearised with *DraII* before PCR amplification, to minimise recovery of this vector in the subsequent cloning reaction.

A Cre reaction was used to catalyse a recombination event between the loxP sites in the PCR product and the $_{PAC}$RC2b vector. The $_{PAC}$RC2b vector (described in Appendix D) is similar to the $_{PAC}$RC1b vector: as above, the recombination method will delete the βgalactosidase cassette, changing the colony colour in the presence of IPTG/Xgal. Different ratios of vector to PCR product were used in the Cre reactions and controls with no PCR product, no $_{PAC}$RC2b vector, or no Cre were included (Table 3.2b – p.124).

2 µl aliquots of these reactions were disk dialysed and electroporated into 30 µl DH10B-u/pSpRecGam competent *E. coli* cells (see Methods - two aliquots each for reactions 1-3, one each for reactions 4 - 6). In addition, a 2 µl aliquot of each reaction was transformed into DH10B-u competent *E. coli* cells as a control, and 100 ng of

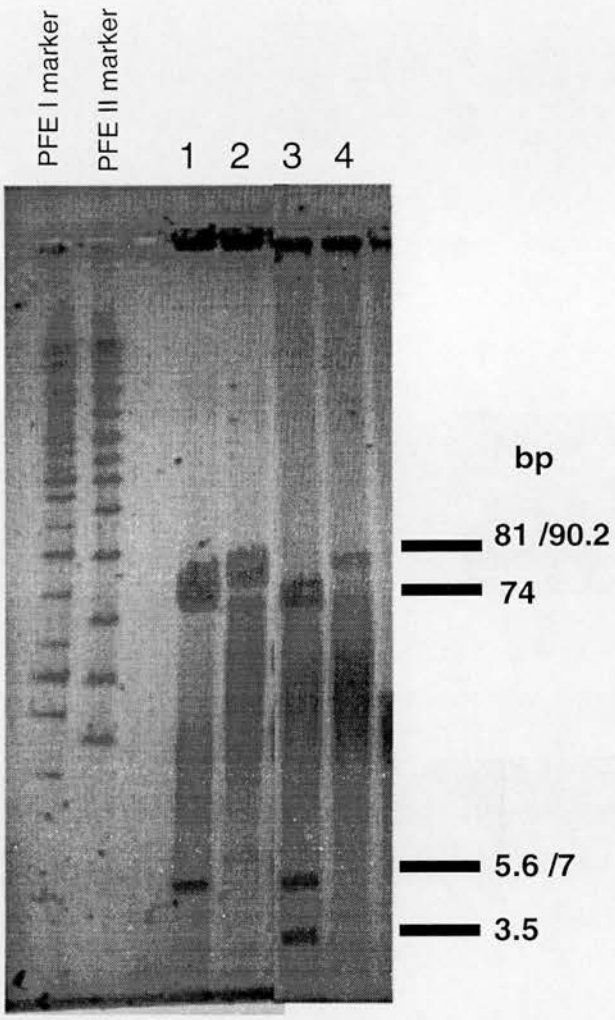


Figure 3.9. Diagnostic digests of $_{PAC}$65kbcfproEGFP and $_{PAC}$RC1b, run down a pulse field gel. The vectors were digested with *SacII* and *PmeI* and run down a pulse field gel. *SacII* digestion of **(1)** $_{PAC}$65kbcfproEGFP produced bands of size 74 and 7 kb and **(3)** $_{PAC}$RC1b produced bands of size 74, 5.6 and 3.5 kb. *PmeI* digestion of **(2)** $_{PAC}$65kbcfproEGFP produced a band of size 81 kb and **(4)** $_{PAC}$RC1b produced a band of size 90.2 kb

$_{PAC}$RC2b was transformed into a 30 µl aliquot of each *E. coli* strain as an additional control. The time constants were between 2.0 and 2.2 ms. The time constant for transforming DH10B-u with ligation #1 was slightly lower (1.4 ms), but this did not seem to have a major effect on the subsequent colony growth. The time constant for transforming ligation #6 into DH10B-u was much lower (0.5 ms) and a popping sound accompanied the electroporation. This indicated that the electroporation was faulty; this could have been due to salt in the electroporation mix, or a bubble in the cuvette. No colonies grew on plates from this transformation.

Cells were plated on L-agar plates containing Kanamycin, IPTG, and Xgal and grown overnight (see Methods). The next day, the plates were scored for growth and colony colour (Table 3.3c – p.127).

The presence of white colonies on the control plates, at a similar frequency to the Cre reaction plates, suggested that most of the white colonies in the Cre reaction plates were not true recombinants. These colonies could have been the result of contamination of the transformation with another Kanamycin resistant vector. Alternatively, they could have been satellite colonies, or βgalactosidase expressing cells that had not yet developed blue colour due to their small size or insufficient Xgal/IPTG in the plate.

Nevertheless, 100 white colonies from the DH10b-u/pSpRecGam transformation plates 1a-c, 2a-c, and 3a-c (see Table 3.3c) were patched onto a new plate for further analysis. Three white colonies from the $_{PAC}$RC2b DH10B-u transformation were also patched for further investigation of these aberrant white colonies.

Of the100 patched colonies from the DH10B-u/pSpRecGam transformation plates 1a–3c, only 56 appeared white after patching; the rest appeared blue. All three 'white' colonies from the $_{PAC}$RC2b plate appeared blue after patching. This suggests that the original colony colour was not a good indicator of βgalactosidase expression.

Colony PCR was performed to identify recombinant clones amongst these colonies. The primers cmvseq and 4R flank the insertion site, and PCR should produce a band of 1.3 kb for the recombinant clone. Because the βgalactosidase cassette that is

removed from the vector during the recombination event is not very large, we found that it was also possible for to amplify a band from the parent vector with these primers. However, this band was larger (4 kb); thus, we were able to distinguish between the parental vector and recombinant clone by the size of the PCR product.

This PCR was performed on the 56 colonies that appeared white and 16 colonies that appeared blue from the patches of DH10B-u/pSpRecGam transformation plates 1a – 3c. In addition, a PCR was performed on $_{PAC}$RC2b DNA as a control. The products were electrophoresed down a 1% TBE gel (Fig. 3.10).

Surprisingly, one of the blue patched colonies (#22) derived from the plates 1a - 3c produced a band of size 1.3 kb, while none of the white patched colonies from these plates produced bands. Amplification of $_{PAC}$RC2b DNA failed to produce the 4 kb band in this experiment; perhaps the conditions were too stringent in this experiment to allow the generation of this large product.

Colony #22 was investigated further. One possible explanation for the colour discrepancy would be if this colony were a composite of the recombinant vector and the parent vector $_{PAC}$RC2b, thus showing blue colour. To test this theory, the patch was restreaked for single colonies. Indeed, both white and blue colonies were derived. Colony PCR showed that the white colonies derived from this streak produced a band of size 1.3 kb for the recombinant, while blue colonies derived from this streak and $_{PAC}$RC2b colonies produced a band of size 4 kb (Fig. 3.11). Thus, a recombinant clone (designated clone #22W) was rescued from the composite colony #22.

A DNA maxiprep of colony #22W was prepared, and restriction analysis was performed to characterise this plasmid. *NotI* and *SacII* digestion of clone #22W produced the correct digest patterns for $_{PAC}$RC2cmvEGFP in contrast to the parent plasmid $_{PAC}$RC2b (Fig. 3.12).

Cmvseq/4R PCR product was amplified from colony #22W DNA, and the product was purified with a Qiaquick PCR purification kit (see Methods). PCR sequencing was performed with a Rhodamine kit, using the primers cmvseq, 2R, 3F, 3R, 4F and

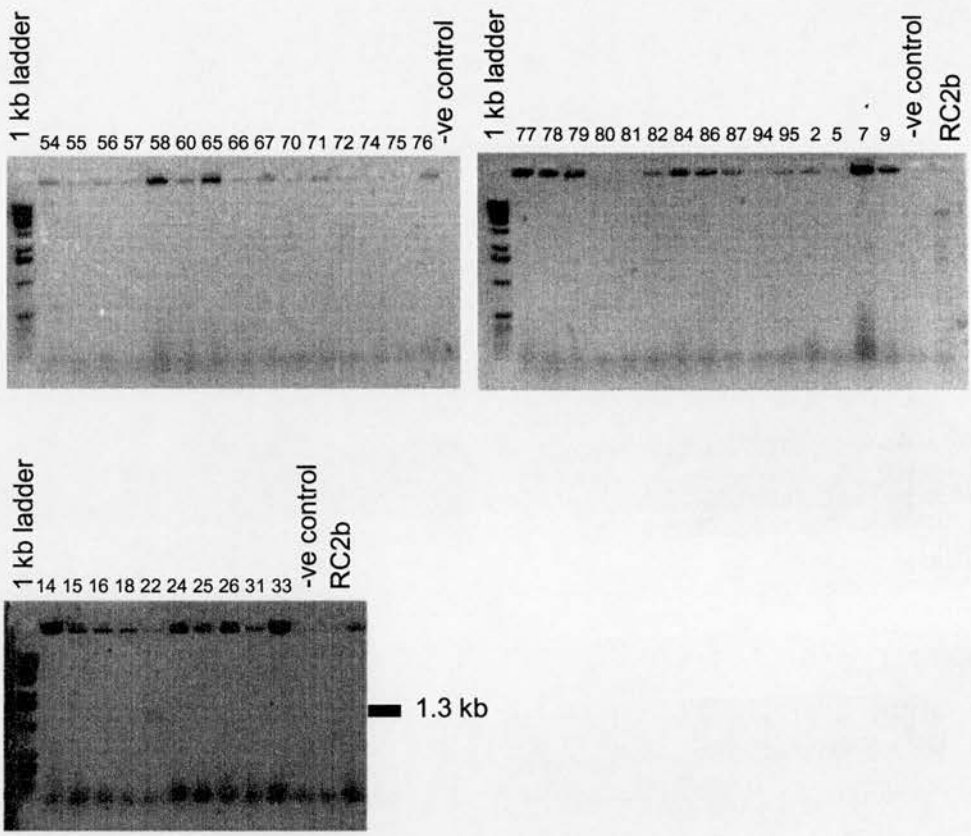


Figure 3.10. cmvseq/4R PCR on $_{PAC}$RC2cmvEGFP colonies. Colony PCR was performed on $_{PAC}$RC2cmvEGFP white colonies 54 - 95, blue colonies 2 - 33, $_{PAC}$RC2b colonies, or a negative control (as labelled), using the primers cmvseq and 4R (Appendix A). None of the white colonies or $_{PAC}$RC2b produced a band of the expected size for either the recombinant or parent plasmid. However, one of the blue colonies, #22, produced a 1.3 kb band - the expected size for the recombinant plasmid.

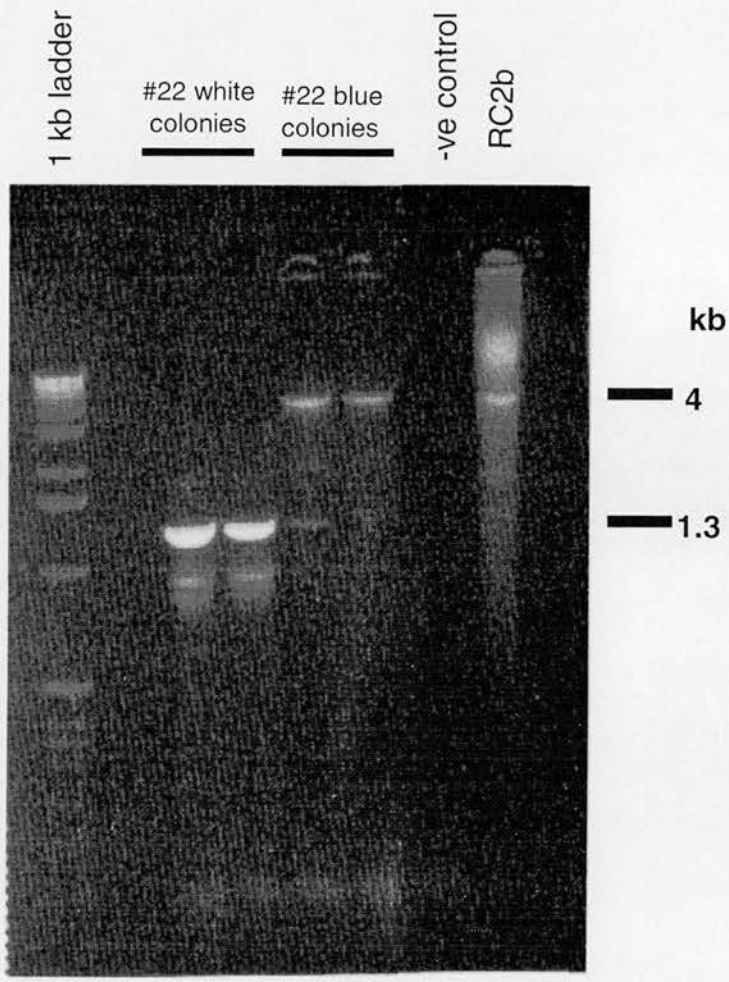


Figure 3.11. cmvseq/4R PCR on $_{PAC}$RC2cmvEGFP clone #22 blue and white colonies . Colony PCR was performed on $_{PAC}$RC2cmvEGFP clone #22 white and blue colonies, a $_{PAC}$RC2b colony,or a negative control (as labelled), using the primers cmvseq and 4R (Appendix A). The two white colonies produced a 1.3 kb band - the expected size for the recombinant plasmid, while the two blue colonies and $_{PAC}$RC2b produced a 4 kb band.

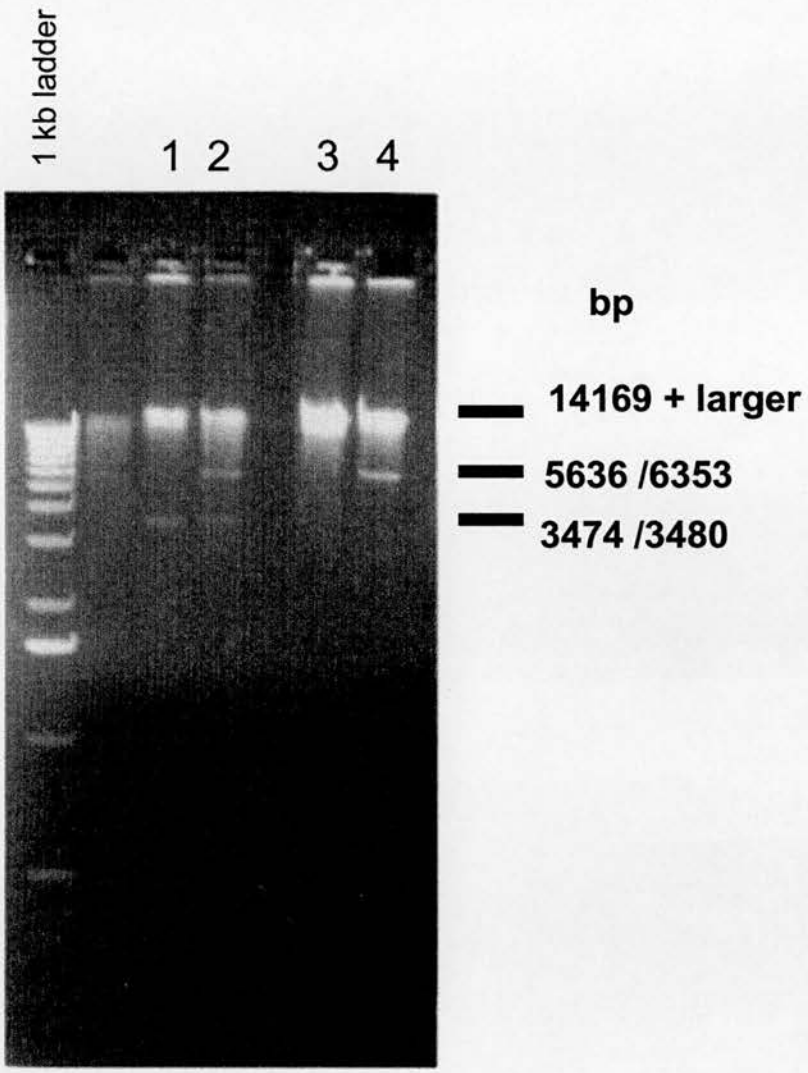


Figure 3.12. Diagnostic digests of $_{PAC}$RC2cmvEGFP clone #22W and $_{PAC}$RC2b. Maxiprep DNA was digested with *NotI* and *SacII* and run down an agarose gel (see Methods). *NotI* digestion of (1) $_{PAC}$RC2b produced two bands: a large band (a predicted doublet of sizes 96624 and 14207 bp), and a small band of size 3474 bp; (3) $_{PAC}$RC2cmvEGFP clone #22W produced a single large band (a predicted doublet of sizes 14169 and 97172 bp). *SacII* digestion of (2) $_{PAC}$RC2b produced three bands: a large band (a predicted doublet of sizes 73956 and 31236 bp) and two smaller bands of sizes 5636 and 3480 bp; (4) $_{PAC}$RC2cmvEGFP clone #22W produced two bands: a large band (a predicted doublet of sizes 73953 and 31035 bp) and a smaller band of size 6353 bp.

4R (see Appendix A). No mutations were found in the recombinant region (the sequencing results are reported in Appendix B3). Thus, clone #22W was named $_{PAC}$RC2cmvEGFP.

Finally, diagnostic restriction digests were performed and run down a pulse-field gel (see Methods) to confirm that the $_{PAC}$RC2cmvEGFP plasmid does not contain any gross rearrangements (Fig. 3.13). *SacII*, *SbfI*, and *NotI* digestion of $_{PAC}$RC2cmvEGFP produced the correct pattern for the recombinant clone, which was similar to the pattern obtained for the $_{PAC}$RC2b parent vector.

3.3.5 $_{PAC}$RC1iresEGFP

$_{PAC}$RC1iresEGFP is a PAC containing the coding sequences of both CFTR and EGFP, separated by an IRES, as in pCFTRiresEGFP. 65 kb of CFTR 5' region is present upstream of the CFTR sequence, to drive expression (see fold-out Appendix E).

$_{PAC}$RC1iresEGFP was made in a similar manner to $_{PAC}$65kbcfproEGFP (Fig. 3.2.5). A region containing a portion of CFTR exon 24, the internal ribosome entry site, and the EGFP gene was amplified from pCFTRiresEGFP, using the primers CFTR5491 and loxP, which introduces the loxP site into the PCR product (Appendix A). The vector pCFTRiresEGFP was linearised with *SnaBI* before amplification, to prevent recovery of this plasmid during the subsequent cloning reaction. A Cre reaction was then used to catalyse recombination between the loxP sites in this PCR product and $_{PAC}$RC1b (see Methods). Vector and insert were mixed in different ratios in the Cre reactions (Table 3.2c – p.124).

2 µl aliquots of these Cre reactions were disk dialysed, and electroporated into DH10B-u/pSpRecGam *E. coli* cells (two aliquots each for reactions 1+ 2 and one aliquot for reactions 3-5). In addition, a 2 µl aliquot of each reaction was transformed into DH10B-u *E. coli* cells. The time constants were between 1.8 and 2.0 ms. The cells were plated on L-agar plates containing Kanamycin, IPTG, and Xgal, and grown overnight (see Methods). The following day, growth and colony colour were scored (Table 3.3d – p.128).

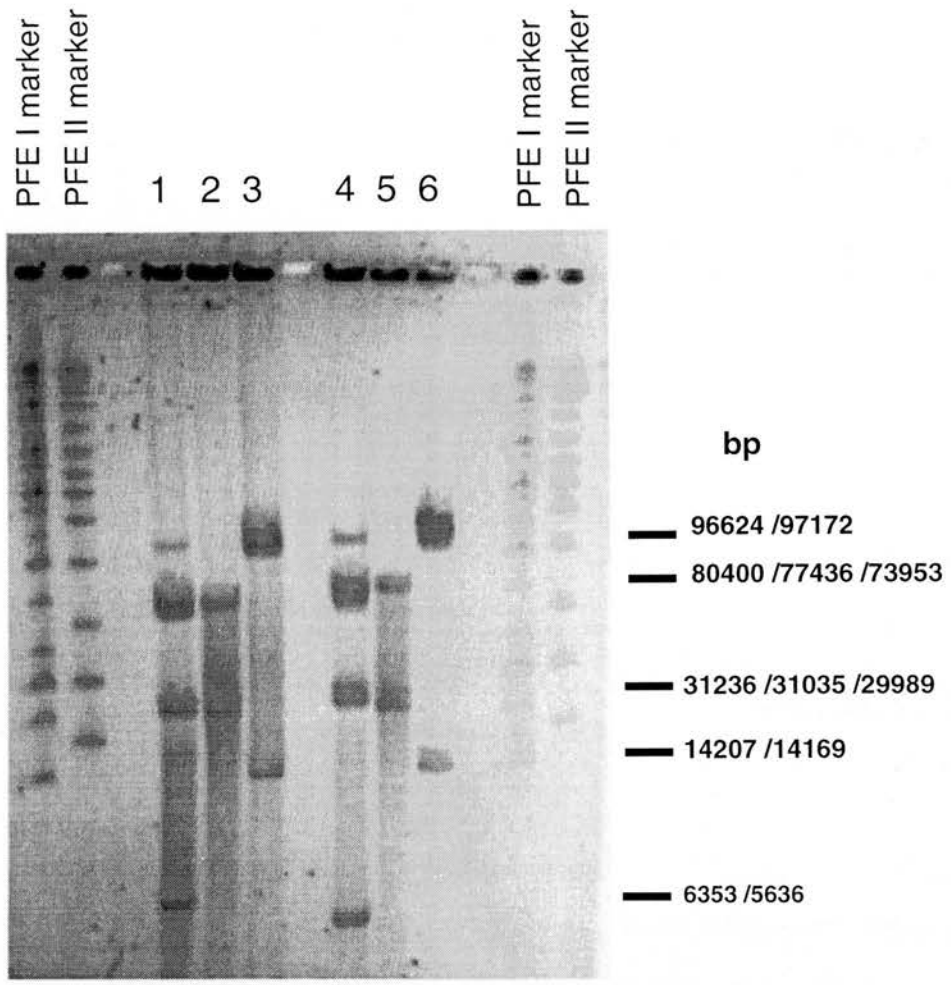


Figure 3.13. Diagnostic digests of $_{PAC}$RC2cmvEGFP and $_{PAC}$RC2b, run down a pulse field gel. The vectors were digested with *SacII, SbfI* and *NotI* and run down a pulse field gel. *SacII* digestion of **(1)** $_{PAC}$RC2cmvEGFP produced three bands (73953, 31035 and 6353 bp) and **(4)** $_{PAC}$RC2b produced three similar bands(73953, 31236 and 5636 bp). *PmeI* digestion of **(2)** $_{PAC}$RC2cmvEGFP produced two bands (77436 and 29989 bp) and **(5)** $_{PAC}$RC2b produced two similar bands (80400 and 29989 bp. *NotI* digestion of **(3)** $_{PAC}$RC2cmvEGFP produced two bands (97172 and 14169 bp) and **(6)** $_{PAC}$RC2b produced two similar bands (96624 and 14207 bp). (N.B.. *SacII* digestion of $_{PAC}$RC2b should also produce a 3480 bp band; *SbfI* digestion of $_{PAC}$RC2cmvEGFP and $_{PAC}$RC2b should also produce 2676 and 1240 bp bands, and *NotI* digestion of $_{PAC}$RC2b should also produce a 3474 bp band: these are too small to be visualised on a pulse field gel).

Unfortunately, Xgal/IPTG βgalactosidase stain was unsuccessful in these plates, and all of the colonies appeared white. Hence, 400 colonies from the DH10B-u/pSpRecGam Cre reaction transformation plates were selected at random and patched onto new plates. 22 of these appeared white after patching (labelled a-v).

Colony PCR was performed on these white colonies using the primers 3F and 4R (see Appendix A) - the product of this PCR spans the 5' breakpoint of insertion. The products were electrophoresed down a 1% TBE gel (Fig. 3.14). Six colonies (a, b, f, h, l and o) produced a band of the correct size (1014 bp), in addition to the positive control $_{PAC}$65kbcfproEGFP (which has the same 5' breakpoint). Negative controls with no template did not produce a band.

A second colony PCR product was performed on 5 of these colonies (a, b, f, h and l) with the primers CF5420 and 4R (see Appendix A); this product spans the entire site of insertion. The products were electrophoresed down a 1% TBE gel (Fig. 3.15). Colonies a, b, h, and l produced an 1874 bp band in this PCR, confirming that they contained the entire insert (in a repeat of this PCR, colony f also produced a band of size 1874 bp - results not shown). The negative controls did not produce a band.

This CF5420/4R PCR product was amplified directly from these colonies; this product was purified with the Qiaquick PCR purification kit and treated with ExoSapIT™ (see Methods). PCR-sequencing was then performed with the primers CF5420, 2R, 3F, 3R, 4F and 4R (Appendix A) using a Rhodamine kit (see Methods).

Every clone sequenced contained the mutation 71082 G → A. We hypothesised that this mutation derived from the original Clontech vector pIRES2EGFP. This mutation would destroy an *ApaLI* restriction site at position 1145 in the pIRES2EGFP vector, leaving a single *ApaLI* site at position 4937. Indeed, *ApaLI* digestion of the pIRES2EGFP vector yielded a single band at ~5308 bp, representing linearised vector (Fig. 3.16). This is consistent with the hypothesis that the mutation derived from the parent vector.

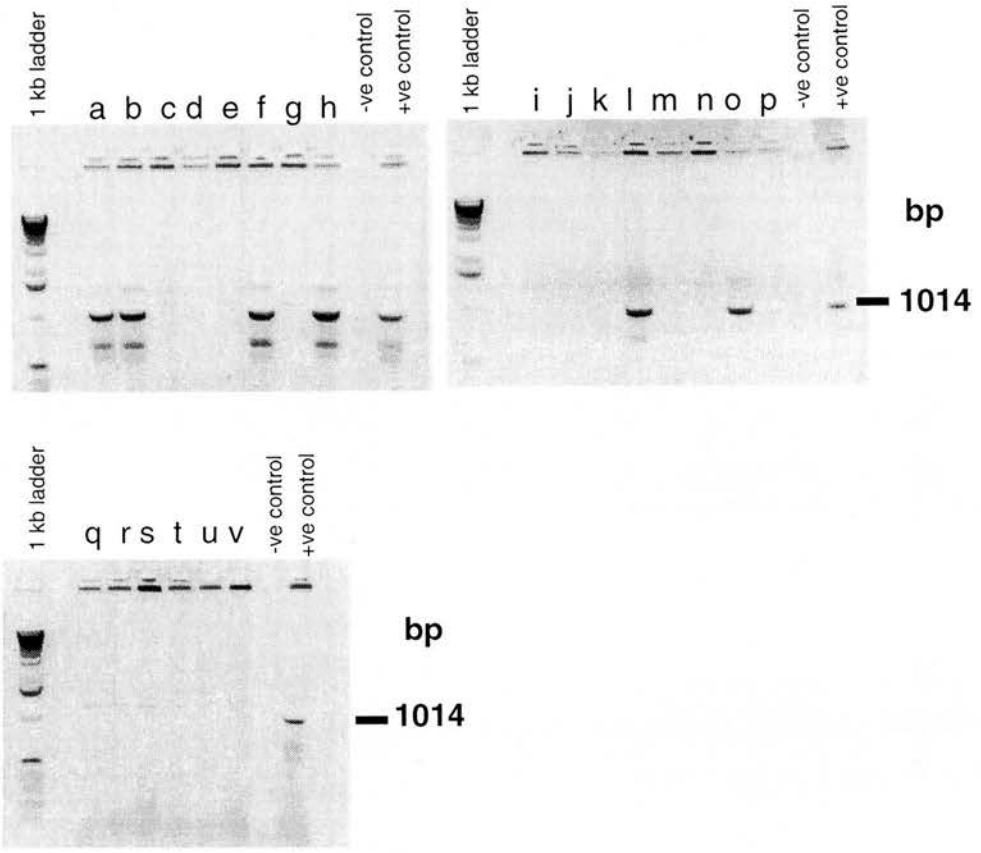


Figure 3.14. 3F/4R PCR on $_{PAC}$RC1iresEGFP colonies.
Colony PCR was performed on $_{PAC}$RC1iresEGFP colonies a-v, or a positive control ($_{PAC}$65kbcfproEGFP) or negative control (as labelled), using the primers 3F and 4R (Appendix A). Colonies a, b, f, h, l and o, and the positive control show a 1014 bp PCR product, indicating that the correct 3' join is present in these colonies.

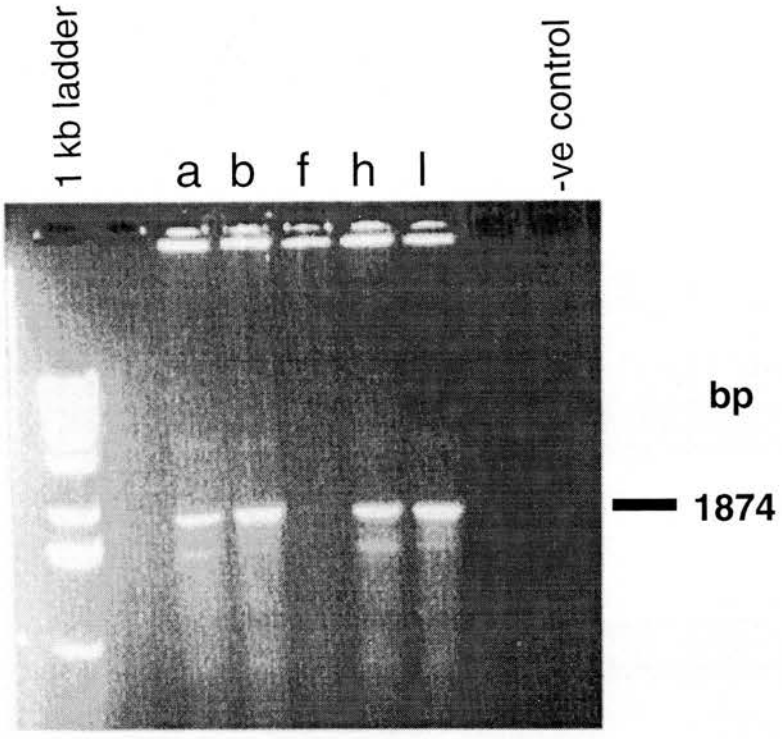


Figure 3.15. cf5420/4R PCR on $_{PAC}$RC1iresEGFP colonies. Colony PCR was performed on $_{PAC}$RC1iresEGFP colonies a-l, or a negative control (as labelled), using the primers cf5420 and 4R (Appendix A). Colonies a,b,h and l show a 1874 bp PCR product, indicating that the correct 5' join is present in these colonies (unfortunately there was no positive control available for this PCR).

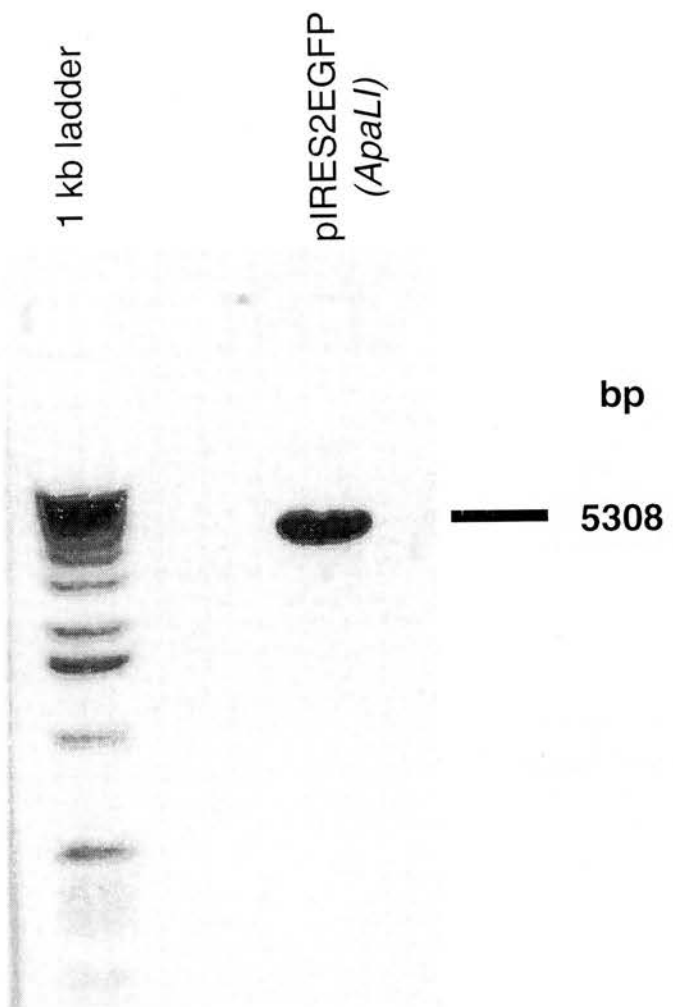


Figure 3.16. *ApaLI* digestion of pIRES2EGFP. pIRES2EGFP DNA was digested with *ApaLI* (as labelled). A single 5308 bp band can be seen (representing linearised vector), suggesting that only one *ApaLI* site is present in the vector.

Sequencing showed that clone l also contained a single base pair deletion in the region between the polyadenylation site of EGFP and the downstream loxP site (72127 ΔT), however this mutation is in an unimportant region, and will not affect expression; otherwise this clone was free from mutations (The sequencing results are reported in Appendix B4). Hence, clone l was named $_{PAC}$RC1iresEGFP.

Restriction digests were performed to further characterise this vector (Fig. 3.17). *NotI* and *SacII* digestion of the $_{PAC}$RC1iresEGFP vector produced the correct pattern of bands in contrast to the $_{PAC}$RC1b parent vector.

In addition, some diagnostic digests were run down a pulse field gel, to confirm that there were no gross rearrangements in the vector (Fig. 3.18). *BsiWI* and *SbfI* digests produced the correct pattern for $_{PAC}$RC1iresEGFP, which was similar to the pattern obtained for the parent vector $_{PAC}$RC1b.

3.3.6 $_{PAC}$RC2iresEGFP

$_{PAC}$RC2iresEGFP is identical to $_{PAC}$RC1iresEGFP with the exception that it contains intron 1 of CFTR in its natural context (see fold-out Appendix E). Thus, these twin vectors can be used to dissect the influence of intron 1 upon expression.

This clone was made in the same way as $_{PAC}$RC1iresEGFP except that the CFTR5491/loxP PCR product (retained from pRC1iresEGFP construction – see above) was recombined with the vector $_{PAC}$RC2b instead of $_{PAC}$RC1b (Fig. 3.2.6). Table 3.2d (p.124) shows the ratios of PCR product to vector used in the Cre reactions.

3 µl aliquots of the Cre reactions were disk dialysed and electroporated into DH10B-u/pSpRecGam cells (two aliquots each for reactions 1 and 2 and one each for reactions 3 and 4). In addition, a 3 µl aliquot of each reaction was transformed into DH10B-u cells as a control. The time constants were generally between 1.9 and 2.0 ms (the time constant of transformation 1a was only 1.6 ms, and fewer colonies grew on plates from this transformation).

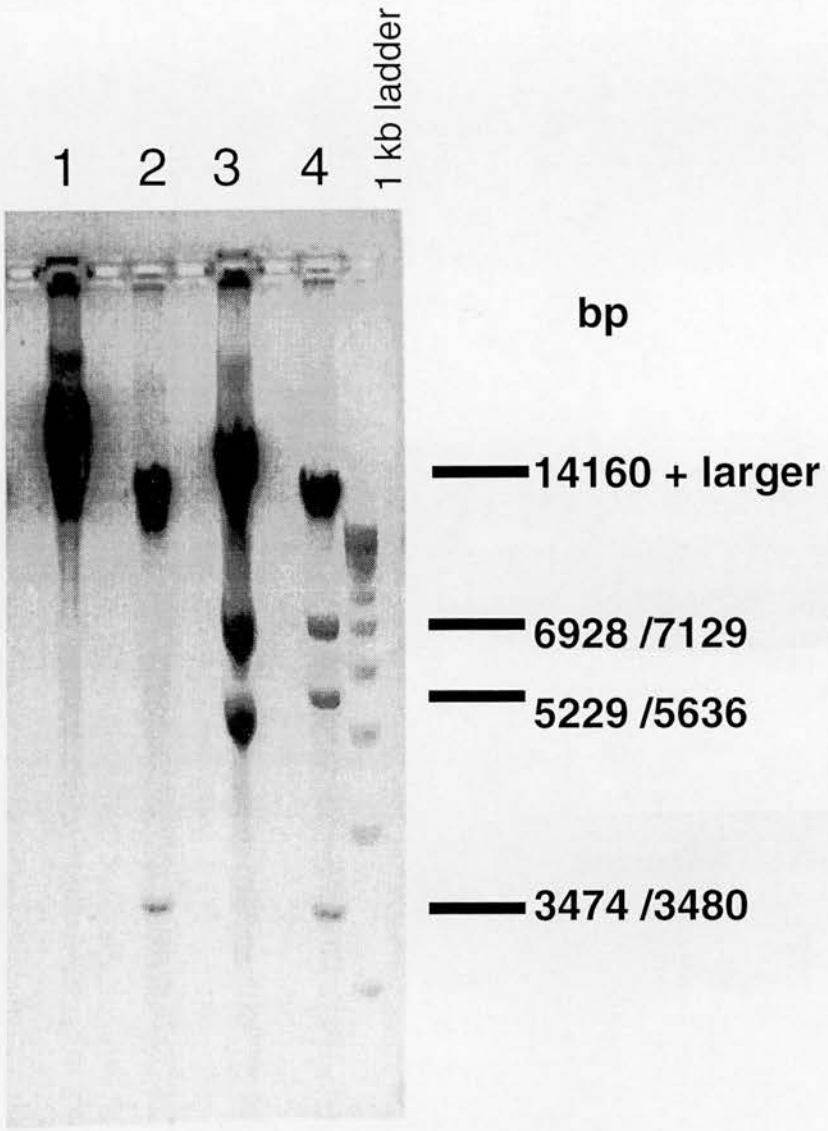


Figure 3.17. Diagnostic digests of $_{PAC}$RC1iresEGFP and $_{PAC}$RC1b. Maxiprep DNA was digested with *NotI* and *SacII* and run down an agarose gel (see Methods). *NotI* digestion of **(1)** $_{PAC}$RC1iresEGFP produced a single large band (a predicted doublet of sizes 71941 and 14160 bp); **(2)** $_{PAC}$RC1b produced two bands: a large band (a predicted doublet of sizes 72517 and 14207 bp) and a smaller band (3474 bp). *SacII* digestion of **(3)** $_{PAC}$RC1iresEGFP produced three bands: a large band (predicted size 73953 bp) and two smaller bands (6928 and 5229 bp); **(4)** $_{PAC}$RC1b produced four bands: a large band (predicted size 73953) and three smaller bands (7129, 5636 and 3480 bp).

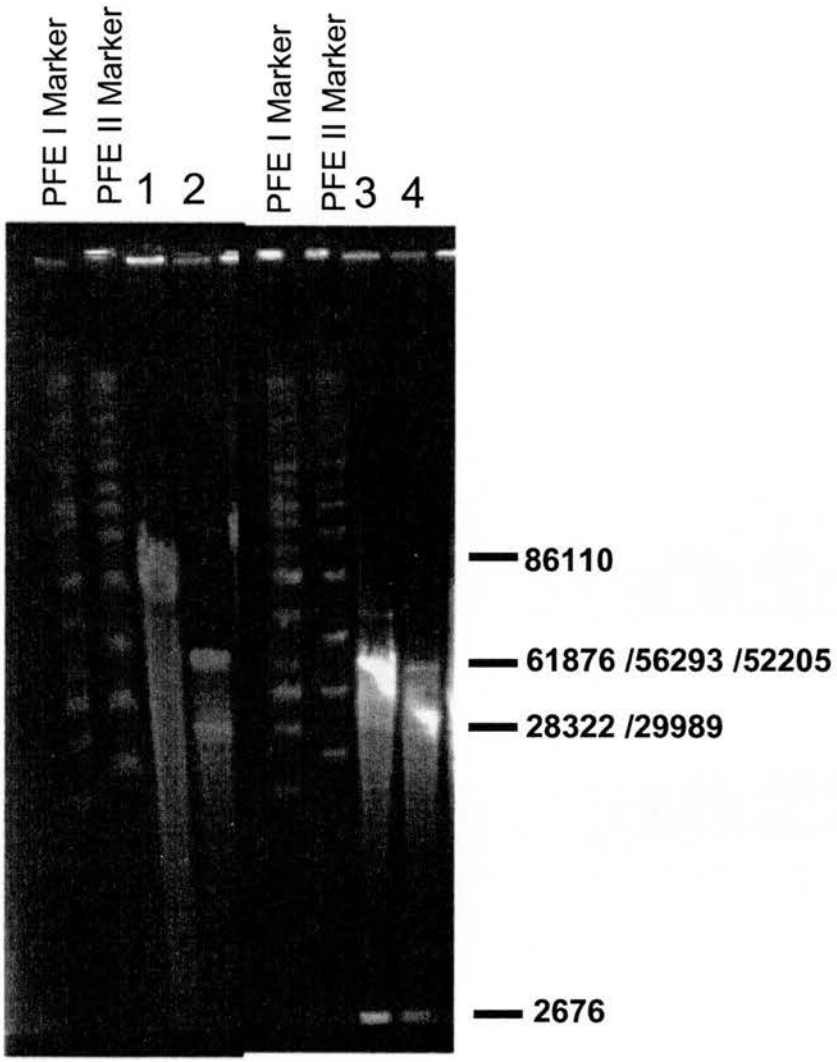


Figure 3.18. Diagnostic digests of $_{PAC}$RC1iresEGFP and $_{PAC}$RC1b, run down a pulse field gel. Maxiprep DNA was digested with *BsiWI* and *SbfI* and run down a pulse field gel (see Methods). *BsiWI* digestion of **(1)** $_{PAC}$RC1iresEGFP produced a single band (86110 bp) **(2)** $_{PAC}$RC1b produced two bands (61876 and 28322 bp). *SbfI* digestion of **(3)** $_{PAC}$RC1iresEGFP produced three bands (52205, 29989 and 2676) **(4)** $_{PAC}$RC1bproduced three similar bands (56293,29989 and 2676 bp).

Cells were plated on L-agar plated containing Kanamycin, IPTG, and Xgal and grown overnight (see Methods). Table 3.3e (p.128) shows the score of colony growth. The 58 white colonies were patched onto a new plate.

Colony PCR was performed on the patched colonies using the primers CF5420 and 4R (see Appendix A - this PCR spans the entire site of insertion). Nine colonies (# 10,11,15,16,17,18,19,22 and 23) produced a band of the correct size for the recombinant clone (1874 bp). Six colonies (#13, 14, 20, 21, 24 and 27) produced a band of ~600 bp size; one might speculate that these were colonies where an aberrant recombination has deleted the βgalactosidase cassette without inserting the ires-EGFP fragment, but these colonies were not investigated further. $_{PAC}$RC1iresEGFP was included as a positive control – this also produced an 1874 bp band, while the negative control did not produce a band (Fig. 3.19).

The CF5420/4R colony PCR product was amplified directly from the colonies, then purified with the Qiaquick PCR purification kit and treated with ExoSapIT™ (see Methods). PCR-sequencing was then performed with the primers CF5420, 2R, 3F, 3R, 4F and 4R (see Apppendix A) using a Rhodamine kit (see Methods). The mutation 95187 G → A was identified in every clone sequenced; this is identical to the mutation seen in the $_{PAC}$RC1iresEGFP clone, and is consistent with the hypothesis that this sequence was derived from the parent vector.

Clone #18 also contained a mutation in the region between the polyadenylation site of EGFP and the downstream loxP site (96232 ΔT). This mutation was identical to the mutation found in the $_{PAC}$RC1iresEGFP vector (which is in an unimportant region of the vector)– a striking coincidence that will be considered in the Discussion. Otherwise, clone #18 was mutation-free (the sequencing results are included in Appendix B5). This clone was named $_{PAC}$RC2iresEGFP.

Finally, restriction digests were performed on $_{PAC}$RC2iresEGFP, to confirm that the vector structure was correct (Fig. 3.20). The *NotI* and *SacII* digestion of $_{PAC}$RC2iresEGFP produced the correct pattern for the recombinant clone, in contrast to the $_{PAC}$RC2b parent vector.

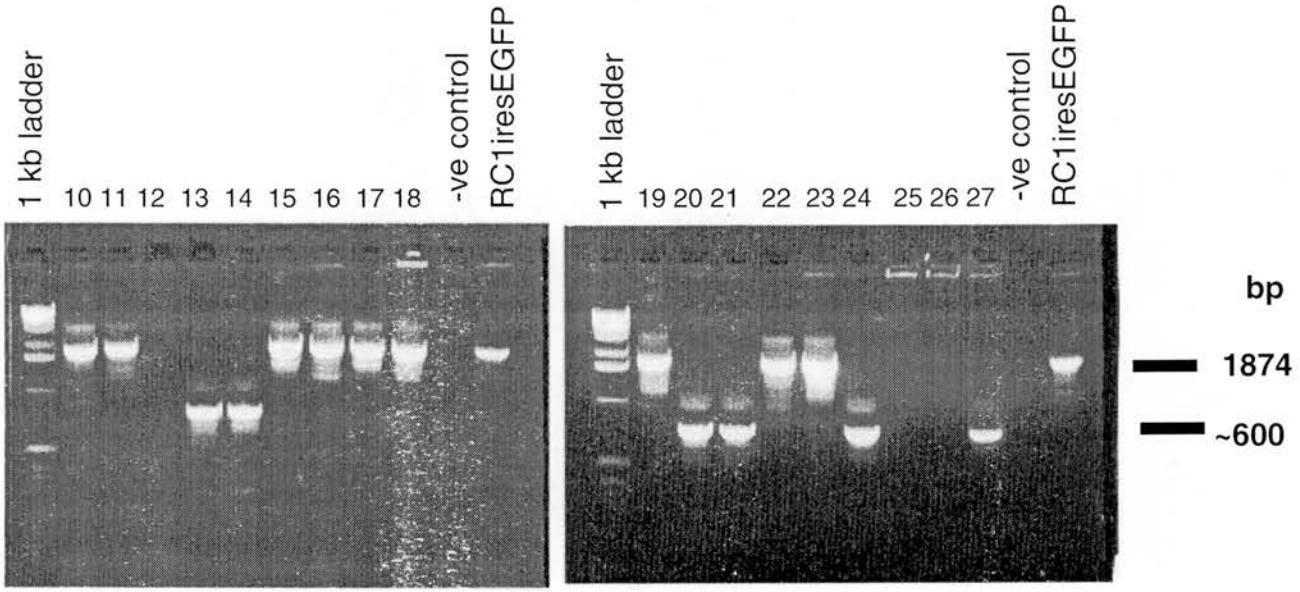


Figure 3.19. cf5420/4R PCR on $_{PAC}$RC2iresEGFP colonies. Colony PCR was performed on $_{PAC}$RC2iresEGFP colonies 10-27, or a positive control ($_{PAC}$RC1iresEGFP) or negative control (as labelled), using the primers cf5420 and 4R (Appendix A). Colonies 10, 11, 15, 16, 17, 18, 19, 22, and 23 and the $_{PAC}$RC1iresEGFP positive control produced a band of size 1874: the correct size for the recombinant clone. Colonies 13, 14, 20, 21, 24 and 27 produced a smaller band of size ~600 bp.

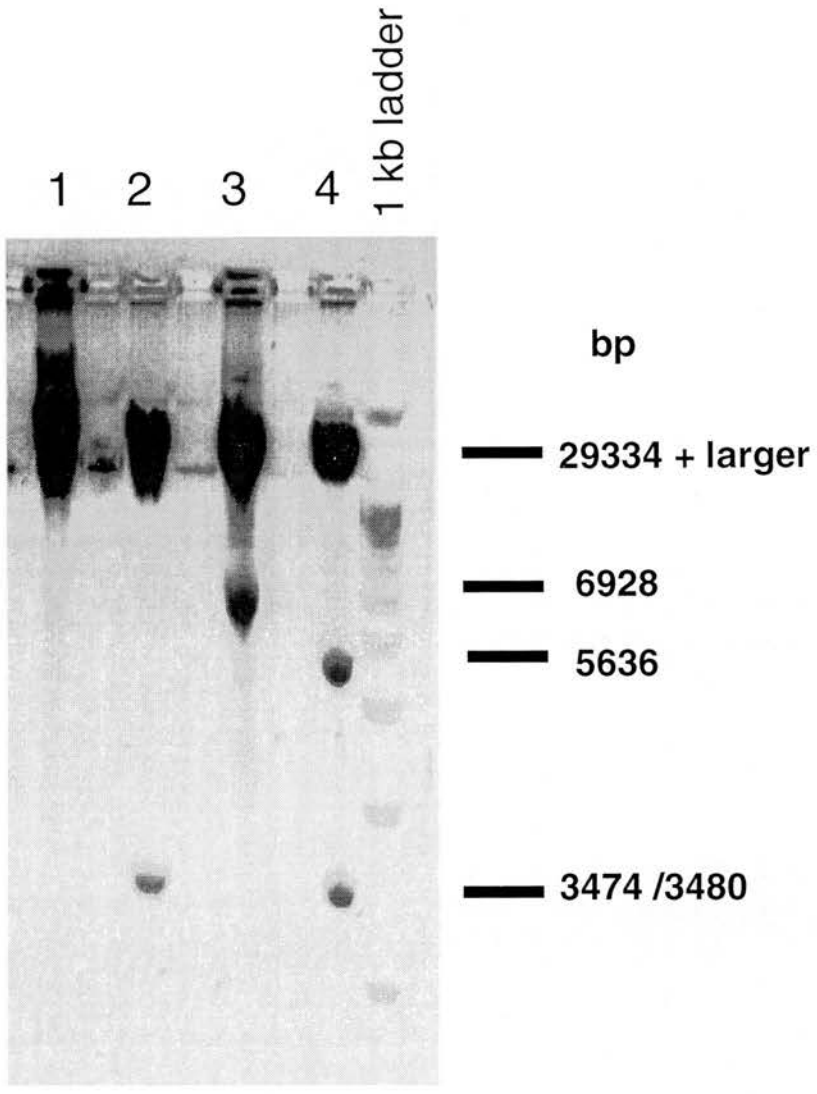


Figure 3.20. Diagnostic digests of $_{PAC}$RC2iresEGFP and $_{PAC}$RC2b. Maxiprep DNA was digested with *NotI* and *SacII* and run down an agarose gel (see Methods). *NotI* digestion of **(1)** $_{PAC}$RC2iresEGFP produced a single large band (a predicted doublet of sizes 96046 and 14169 bp); **(2)** $_{PAC}$RC2b produced two bands: a large band (a predicted doublet of sizes 96624 and 14207 bp) and a smaller band (3474 bp). *SacII* digestion of **(3)** $_{PAC}$RC2iresEGFP produced two bands: a large band (a predicted doublet of sizes 73953 and 29334 bp) and a smaller band (6928 bp); **(4)** $_{PAC}$RC2b produced three bands: a large band (a predicted doublet of sizes 73953 and 31236) and two smaller bands (5636 and 3480 bp).

In addition, some $_{PAC}$RC2iresEGFP diagnostic digests were run down a pulse field gel to confirm that there were no gross rearrangements in the vector (Fig. 3.21). *BsiWI* and *SbfI* digestion of $_{PAC}$RC2iresEGFP produced the correct pattern for the recombinant vector, which was similar to the $_{PAC}$RC2b parent vector.

3.4 Discussion

The successful creation of the vectors $_{PAC}$65kbcfproEGFP, $_{PAC}$RC2cmvEGFP, $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP validates some theoretical concepts of vector design. Firstly, the creation of these vectors demonstrates that it is possible to use a PCR product in the established double recombination method (Boyd et al, 1999b). Secondly, it has proved possible to incorporate a loxP site onto the end of a PCR product by primer-directed mutagenesis.

Colony PCR is a powerful tool for identifying recombinant clones generated by the double-recombination method. A large number of white colonies were generated during the production of $_{PAC}$RC2cmvEGFP (56 white colonies) that were not true recombinants. In addition, there were many repeats of the cloning experiments which generated no recombinant clones whatsoever, despite the presence of many white colonies (data not shown). Because true recombinants were very rare in some cloning reactions, it would have been laborious to identify these clones by making crude DNA minipreps and performing restriction analysis.

During the analysis of $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP vectors, PCR sequencing was successfully performed using the product of a colony PCR as template. This is much less laborious (and expensive) than performing a DNA maxiprep on every colony prior to PCR-sequencing. Bacterial residue is removed from the colony PCR during the Qiaquick purification and ExoSapIT™ treatment to a sufficient degree to facilitate subsequent sequencing reactions.

The $_{PAC}$RC2cmvEGFP recombinant clone was rescued from a mixed colony. The blue colour derived from the parent plasmid ($_{PAC}$RC1b) in colony #22 masked the presence of the recombinant clone, which would otherwise produce white bacteria.

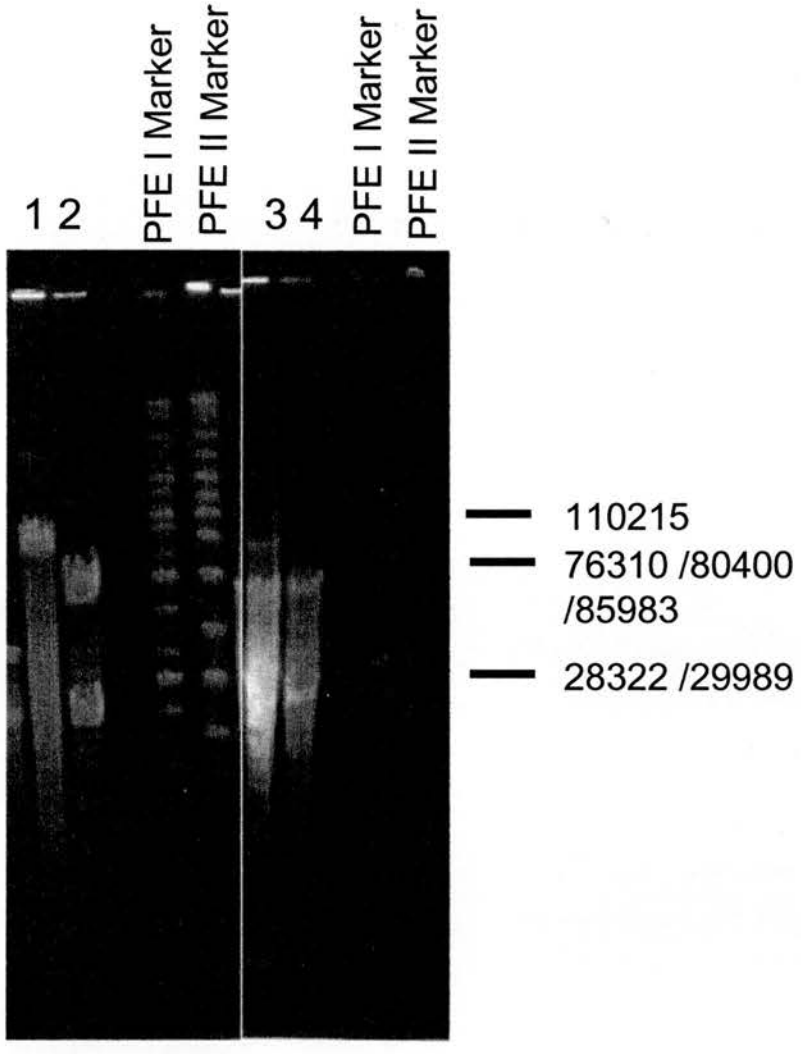


Figure 3.21. Diagnostic digests of $_{PAC}$RC2iresEGFP and $_{PAC}$RC2b, run down a pulse field gel. Maxiprep DNA was digested with *BsiWI* and *SbfI* and run down a pulse field gel (see Methods). *BsiWI* digestion of **(1)** $_{PAC}$RC2iresEGFP produced a single band (110215 bp); **(2)** $_{PAC}$RC2b produced two bands (85983 and 28322 bp). *SbfI* digestion of **(3)** $_{PAC}$RC2iresEGFP produced two bands (76310 and 29989); **(4)** $_{PAC}$RC2b produced two similar bands (80400 and 29989). *SbfI* digestion of both $_{PAC}$RC2iresEGFP and $_{PAC}$RC2b also produced a band of size 2676 bp, which is too faint to see on the scanned photo.

Through the sequencing of daughter clones, a probable variation from the published sequence was uncovered for the vector pIRES2EGFP. This variation would destroy one of two predicted *ApaLI* restriction sites in the pIRES2EGFP vector. *ApaLI* restriction linearised this vector, rather than producing two fragments, supporting the hypothesis that the sequence variation derived from the Clontech vector. In other words, the reported sequence for the pIRES2EGFP plasmid appears to be incorrect.

The $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP clones contained an identical mutation (ΔT) between the polyadenylation site of EGFP and the loxP site. This region was derived from the loxP primer. Several clones that were sequenced (but disqualified due to other mutations) did not contain the ΔT mutation. However, the presence of this identical mutation in both vectors is an uncanny coincidence, and may still reflect an aberrant synthesis of the loxP primer, such that a proportion of primers contain this divergent sequence. Fortunately, this mutation does not alter a functional region of the vector and is therefore inconsequential.

Different ratios of vector to PCR product to insert were included in the Cre reactions in an attempt to identify the optimum ratio for generating recombinant clones. For example, in the construction of $_{PAC}$65kbcfproEGFP, Table 3.3b (p.126) correlates the patched colony number with the transformation from which it was derived; Fig. 3.6 shows the results of colony PCR on these patched colonies. By analysing this case in point, we can consider whether the insert to vector ratio determines the success of generating recombinant clones.

- The three DH10B-u/pSpRecGam electroporations from Cre reaction 1 produced three colonies that initially appeared white; only one of these proved to be a true white colony after patching. This colony was shown to be a true recombinant in the colony PCR.

- The two electroporations from reaction 2 produced 20 colonies that initially appeared white, although only 13 of these appeared white after patching. Two of these colonies were shown to be true recombinants by colony PCR.

- The three electroporations from reaction 3 produced 54 colonies that initially appeared white. After patching, only 35 of these colonies appeared white. Sixteen of these colonies were shown to be true recombinants by colony PCR.

Thus, the average number of white colonies generated per plate was 0.11 with reaction 1 (1:5 ratio of insert: vector), 3.33 with reaction 2 (1:1 ratio), and 3.89 with reaction 3 (5:1 ratio). It seems that there is a correlation between the insert to vector ratio and the number of white colonies produced. This trend is confirmed by the $_{PAC}$RC2iresEGFP results (see Table 3.3e – p.128) where reaction 1 (1:1 ratio of insert to vector) produced an average of 1.33 white colonies per plate, while reaction 2 (5:1 ratio) produced an average of 8.33 white colonies per plate.

One may also consider the proportion of white colonies that turned out to be true recombinants instead of aberrant white colonies (as determined by colony PCR) and test whether this is affected by the ratio of insert to vector. When we analyse the population of white clones for $_{PAC}$65kbcfproEGFP, we see that 1/1 (frequency = 1) white colonies is a true recombinant with reaction 1 (1:5 ratio of insert: vector), 2/13 (frequency = 0.154) white clones are true recombinants with reaction 2 (1:1 ratio), and 16/35 (frequency = 0.457) white clones are true recombinants with reaction 3 (5:1 ratio). Thus, there is no recognisable trend in the frequency of true recombinants vs. aberrant white colonies, in relation to the insert to vector ratio. Taking both of these factors into account, the largest number of true recombinant clones was derived by using a high ratio of insert to vector.

To sum up, six vectors were created, coupling CFTR genomic context elements to the EGFP reporter gene. These constructs can be used in tranfection studies to investigate the influence of proximal and distal CFTR 5' region and intron 1 elements upon EGFP expression. Four of these vectors were created by a recombinogenic method utilising both Cre/loxP and homologous recombination. Two innovations of vector design were validated through the creation of these constructs: (1) the use of a PCR-generated fragment in the recombinogenic method, and (2) the incorporation of a loxP site onto the end of the PCR product by primer-directed mutagenesis.

Table 3.1. Ligations. This table describes the amount of insert and vector used in the p1kbcfproEGFP ligation reaction:

Reaction #	1	2	3
Digested PCR product	75 ng	50 ng	30 ng
Digested pEGFP-1	62.5 ng	62.5 ng	62.5 ng
10X buffer	1 µl	1 µl	1 µl
Ligase	1 µl	1 µl	1 µl
H2O	→ 10 µl	→ 10 µl	→ 10 µl

Table 3.2. CRE reactions. This table describes the amounts of insert and vector used in the Cre reactions during the construction of **(a)** $_{PAC}$65kbcfproEGFP, **(b)** $_{PAC}$RC2cmvEGFP, **(c)** $_{PAC}$RC1iresEGFP, and **(d)** $_{PAC}$RC2iresEGFP.

(a) $_{PAC}$65kbcfproEGFP

Reaction#	1	2	3	4	5
RC1b	500ng	500ng	100ng	500ng	500ng
PCR product	100 ng	500 ng	500 ng	-	500 ng
Buffer M	1 µl	1 µl	1 µl	1 µl	1 µl
CRE	0.5 µl	0.5 µl	0.5 µl	0.5 µl	-
H2O	→ 10 µl	→ 10 µl	→ 10 µl	→ 10 µl	→ 10 µl

(b) $_{PAC}$RC2cmvEGFP

	1	2	3	4	5	6
RC2b	500 ng	500 ng	100 ng	500 ng	500 ng	-
cmv1/loxP PCR	500 ng	2.5 µg	500 ng	500 ng	-	500 ng
Buffer M	1.5 µl	1.5 µl	1.5 µl	1.5 µl	1.5 µl	1.5 µl
CRE	0.5 µl	0.5 µl	0.5 µl	-	0.5 µl	0.5 µl
H_2O	→ 15 µl	→ 15 µl	→ 15 µl	→ 15 µl	→ 15 µl	→ 15 µl

(c) $_{PAC}$RC1iresEGFP

	1	2	3	4	5
Rc1b	100 ng	100 ng	100 ng	100 ng	-
PCR product	100 ng	500 ng	-	100 ng	100 ng
Buffer M	1 µl	1 µl	1 µl	1 µl	1 µl
CRE	0.5 µl	0.5 µl	0.5 µl	-	0.5 µl
H_2O	→ 10 µl	→ 10 µl	→ 10 µl	→ 10 µl	→ 10 µl

(d) $_{PAC}$RC2iresEGFP

	1	2	3	4	5
Rc2b	500 ng	500 ng	500 ng	500 ng	-
PCR product	500 ng	2.5 µg	-	500 ng	500 ng
Buffer M	1 µl	1 µl	1 µl	1 µl	1 µl
CRE	0.5 µl	0.5 µl	0.5 µl	-	0.5 µl
H_2O	→ 10 µl	→ 10 µl	→ 10 µl	→ 10 µl	→ 10 µl

Table 3.3. Scores of colony growth. This table shows the score of colony growth from transformations of **(a)** p1kbcfproEGFP, **(b)** $_{PAC}$65kbcfproEGFP, **(c)** $_{PAC}$RC2cmvEGFP, **(d)** $_{PAC}$RC1iresEGFP, and **(e)** $_{PAC}$RC2iresEGFP. Where colony colour was scored. B=blue, W=white.

(a) p1kbcfproEGFP

Ligation #	plate	Volume plated	# colonies
1	A	100	19
	B	200	34
	C	200	43
2	A	100	14
	B	200	47
	C	200	34
3	A	100	11
	B	200	19
	C	200	18

(b) $_{PAC}$65kbcfproEGFP

E.coli strain	Transformation	volume plated	Plate 1	Plate 2	Plate 3	Patched colony #s
DH10B-u/ pSpRecGam	1a	200 µl	~1,000B 0W	~900B 1W	~1,000B 0W	#1
	1b	200 µl	~900B 1W	~900B 0W	~800B 0W	#2
	1c	200 µl	~1,000B 0W	~900B 0W	~800B 1W	#3
	2a	200 µl	~2,800B 5W	~3,000B 4W	~1,000B 3W	#4-15
	2b	200 µl	~1,800B 1W	~2,000B 6W	~1,200B 1W	#16-23
	3a	200 µl	~600B 12W	~550B 9W	~200B 8W	#24-52
	3b	200 µl	~250B 11W	~300B 4W	~120B 2W	#53-69
	3c	200 µl	~320B 4W	~750B 2W	~84B 2W	#70-77
	4	100 µl	~300B 0W			
	5	100 µl	~600B 0W			
DH10B-u	1	100 µl	~4,000B 0W			
	2	100 µl	~3,500B 0W			
	3	100 µl	~3,000B 0W			
	4	100 µl	~5,500B 0W			
	5	100 µl	~6,000B 0W			

(c) $_{PAC}$RC2cmvEGFP

E.coli strain	Transformation	Volume plated	Plate 1	Plate 2	Plate 3
DH10B-u/ pSpRecGam	1a	200	~5,000B, 11W	~6,000B, 7W	~5,500B, 4W
	1b	200	~5,000B, 10W	~5,500B, 10W	~5,000B, 9W
	1c	200	~4,000B, 3W	~5,500B, 5W	~2,000B, 7W
	2a	200	~2,000B, 3W	~2,500B, 3W	~1,500B, 12W
	2b	200	~3,000B, 12W	~4,000B, 13W	~6,000B 10W
	2c	200	~3,000B, 11W	~4,000B, 12W	~3,000B, 16W
	3a	200	~3,000B, 3W	~2,000B, 2W	~1,500B, 5W
	3b	200	~3,500B, 8W	~4,000B, 5W	~4,500B, 7W
	3c	200	~4,000B, 4W	~4,500B, 6W	~5,000B, 8W
	4	50	~2,000B 5W		
	5	50	~300B 0W		
	6	50	0-		
	RC2b	50	~4,000B, 12W		
DH10B-u	1	50	~2,000B 3W		
	2	50	~4,000B, 3W		
	3	50	~2,500B, 0W		
	4	50	~2,500B 3W		
	5	50	~3,000B 2W		
	6	50	0-		
	RC2b	50	Lawn blue, 5W		

(d) $_{PAC}$RC1iresEGFP

E. coli strain	reaction	Volume Plated	Plate 1	Plate 2	Plate 3
DH10B-u/ pSpRecGam	1a	200	~4,000	~5,000	~5,000
	1b	200	~7,000	~7,000	~1,000
	2a	200	~1,500	~3,000	~900
	2b	200	~5,000	~6,000	~2,000
	3	50	~800	-	-
	4	50	~1,500	-	-
	5	50	0	-	-
DH10B-u	1	50	34	-	-
	2	50	0	-	-
	3	50	409	-	-
	4	50	31	-	-
	5	50	0	-	-

(e) $_{PAC}$RC2iresEGFP

E. coli strain	reaction	Volume Plated	Plate 1	Plate 2	Plate 3
DH10B-u/ pSpRecGam	1a	200	71B, 0W	69B, 0W	74B, 0W
	1b	200	488B, 2W	521B, 4W	384B, 2W
	2a	200	561B, 7W	482B, 14W	533B, 14W
	2b	200	328B, 4W	312B, 4W	230B, 7W
	3	50	~400B, 0W		
	4	50	~330B, 0W		
	1	50	~1,800B, 0W		
	2	50	~2,500B, 0W		
	3	50	218B, 0W		
	4	50	2,500B, 0W		

Chapter 4: Optimising transfection efficiency

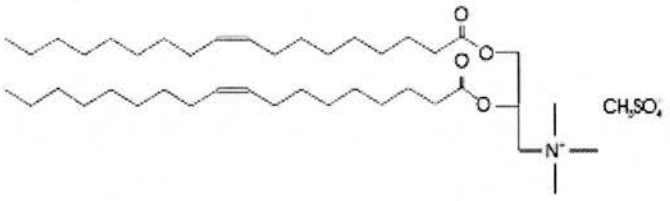
4.1 Introduction

In order to test the effects of EGFP reporter plasmids in permanent cell lines, it was necessary to establish a reliable and efficient method of delivering DNA to these cells. While naked DNA may be sufficient to accomplish a low level of transfection, the addition of transfection reagents has the potential improve delivery dramatically. For example, transfection reagents may enhance delivery by condensing the DNA, targeting a complex for endocytosis, protecting DNA from endosome degradation, or by aiding DNA transport into the nucleus. To optimise transfection efficiency, several reagents were compared:

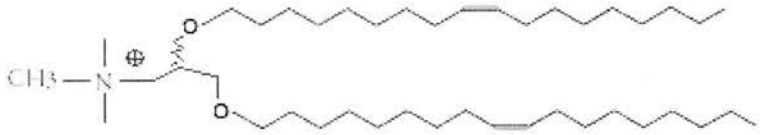
- **DOTAP:** The DOTAP reagent is a liposome formulation of the cationic lipid DOTAP (N-(1-(2,3-Dioleoyloxy)propyl)-N,N,N-trimethylammonium methyl-sulfate) (Fig. 4.1). 12 µl DOTAP was combined with 1 µg DNA in Optimem-1, according to the optimised protocol (McLachlan et al, 1995, see Methods). The reagent is supplied as an aqueous dispersion at 1 mg/ml. Mixing DOTAP with DNA results in the spontaneous formation of stable complexes that are efficiently endocytosed into eukaryotic cells, with low cytotoxicity. This reagent can be obtained from Roche Diagnostics (Roche Diagnostics Ltd., Bell Lane, Lewes, East Sussex, BN7 1LG, UK, http://www.roche-applied-science.com/). DOTAP was used as a 'gold standard' for transfection studies, as this reagent has been extensively employed in the past for *in vitro* and *in vivo* transfection, and has been tested in a clinical CF gene therapy trial (McLachlan et al, 1995; McLachlan et al, 1996; Porteous et al, 1997). However, reagents developed more recently may produce higher levels of transfection.

- **Saint-Mix™:** Saint-Mix™ is a synthetic amphiphilic (non-liposomal) delivery system, developed by Saint-B.V (now Synvolux Therapeutics). Saint-18™ (1-methyl-4-(cis-9-dioleyl)methyl-pyridinium-chloride) (Fig. 4.1) is a member of the pyridinium surfectant family, which effectively delivers DNA, RNA and protein to cells, and has a favourable cytotoxicity profile. The Saint-Mix™ reagent is comprised of Saint-18™

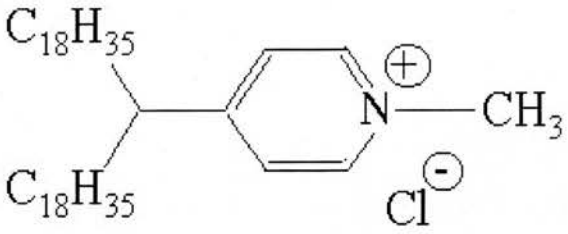
(a) DOTAP



(b) DOTMA



(c) Saint-18



(d) DOPE

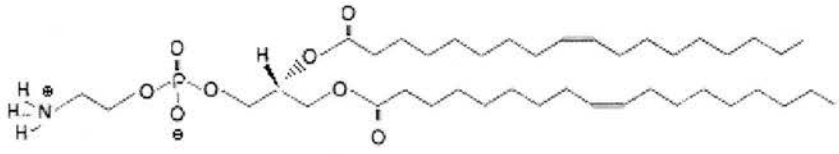


Figure 4.1: Chemical structures of transfection reagents. This figure shows the chemical structures of: **(a) DOTAP** (N-(1-(2,3-Dioleoyloxy)propyl)-N,N,N-trimethylammonium methyl-sulfate) **(b) DOTMA** (N-(1-(2,3-dioleyloxy)propyl)-n,n,n-trimethylammonium chloride) **(c) Saint-18** (1-methyl-4-(cis-9-dioleyl)methyl-pyridinium-chloride) and **(d) DOPE** (Dioleoyl-phosphatidylethanolamine).

combined with DOPE (Dioleoyl-phosphatidylethanolamine) (Fig. 4.1) in a 1:1 ratio (weight:weight). Saint-Mix is supplied at 0.75 mM in an aqueous solution. For optimum transfection efficiency, the company recommends complexing 20 µl of Saint-Mix with 1 µg of DNA (see Methods). This reagent, along with further information, can be obtained from Synvolux Therapeutics (Synvolux Therapeutics, 'Meditech Center', L.J. Zielstraweg 1, 9713 GX Groningen, The Netherlands, http://www.synvolux.nl/).

- **Polyethylenimine (PEI):** PEI has been used to promote transfection *in vitro* and *in vivo* in several studies (Baker et al, 1997, Bragonzi et al, 1999, Boussif et al, 1995, Abdallah et al, 1996). PEI acts as a proton sponge across a wide range of pH values, providing substantial buffering capacity: neutralising negative charge, and condensing DNA, thus enhancing transfection. In the PEI method used here, DNA is pre-condensed with a 22KDa PEI, then condensed with a 25KDa PEI (Thorpe et al, 2002; this method was described in a personal communication from Matt Cotton). The 22KDa and 25 KDa PEIs were a kind gift from Dr. Matt Cotton, Institute for Molecular Pathology, Vienna, Austria.

- **LID:** The LID method aims to enhance gene transfer by adding a targeting ligand to the lipid-DNA complex. The integrin receptor is a particularly attractive target for this purpose; egg-sperm fusion is integrin-mediated and several viruses (e.g. the foot and mouth virus) bind to integrin receptors to promote cell entry. The LID method combines 0.75 µl of the liposome Lipofectin with 4 µg of the integrin targeting peptide P6 and 1 µg of DNA to form a transfection complex (Hart et al, 1995 and 1998, see methods). The Lipofectin reagent is comprised of DOTMA (N-(1-(2,3-dioleyloxy)propyl)-n,n,n-trimethylammonium chloride) (see Figure 4.1) and DOPE in a 1:1 ratio (weight:weight), and can be obtained from Invitrogen (Invitrogen Ltd., 3 Fountain Drive, Inchinnan Business Park,Paisley PA4 9Rf, UK, http://www.invitrogen.co.uk). The P6 peptide was a kind gift from Dr. Steve Hart, Institute of Child Health, London. P6 is comprised of the sequence (K_{16})GACRRETAWACG; this peptide acts as a ligand for the $\alpha5\beta1$ integrin,

which is expressed by epithelial cells of the lung. Theoretically, the P6 peptide will bind to this integrin receptor on the cell surface and the complex will subsequently be taken up into the cell by receptor-mediated endocytosis. LID has been successful in pulmonary gene transfer following lung instillation delivery to the rat (Jenkins et al, 2000). Furthermore, in a slightly different method, cationic nanoparticles incorporating an integrin-targeting ligand sucessfully targeted tumour vasculature *in vivo*, resulting in a substantial level of tissue-specific gene transfer (Hood et al, 2002).

- **SID:** A new transfection protocol was developed for this work, combining the Saint-Mix™ reagent with the integrin-targeting peptide P6: this was named the SID method.

As the project progressed, different assays were used to measure transfection efficiency. At the start of the project, the pCMVβ plasmid was transfected into cells, and βgalactosidase enzyme activity was measured to determine transfection efficiency. Initially, an ELISA was used to detect βgalactosidase expression. This method was replaced with a βgalactosidase luminometer assay, which was quicker and easier, allowing more samples to be processed in parallel.

In keeping with the aims of this project, transfection studies quickly progressed to utilise constructs expressing the EGFP reporter gene. One of the advantages of using the EGFP reporter gene is the capacity to detect expression by fluorescence activated cell sorting (FACS).

FACS is a powerful method. As each cell passes through a cytometer chamber, its individual fluorescence is measured. When data from the entire population is assembled, it is possible to determine (1) the number of cells expressing EGFP, that is, the number of cells with a green fluorescence above background levels and (2) the average fluorescence of this population, which is directly proportional to the level of EGFP expression (Subramanian and Srienc, 1996). With the FACS detection system, one can distinguish between a few cells expressing high levels of protein, and many cells expressing low levels of protein. To obtain such a measure of βgalactosidase expression, it would be necessary to employ two assays: a staining method, to

quantify the number of cells expressing the protein, and a total protein assay, to measure the amount of protein produced.

Initially, a traditional histogram method was used to analyse the FACS output. Later in the project, this was replaced by a dotplot analysis, which was shown to provide improved discrimination between genuine and background fluorescence.

Because of the high level of variability in transfection studies, every experiment was performed on at least two, and usually three separate occasions. Although the absolute levels of transfection often varied from day to day for unknown reasons, the trends in relative efficiency between transfection reagents were consistent between experiments. The results of typical experiments are shown to illustrate each example.

4.2 Results

4.2.1 FACS analysis

Historically, FACS output has been analysed by creating a logarithmic histogram of FL1-H (green) fluorescence; in other words by analysing the green fluorescence intensity of each cell. Untransfected cells produce a single peak with a normal distribution at low fluorescence values. This is termed background fluorescence. Fig. 4.2 shows some FL1-H logarithmic histograms of COS-7 cells with different treatments. In the untransfected sample, there is a single peak on the histogram, centered at ~$10^{0.2}$ relative fluorescence units.

When cells are transfected with a construct expressing EGFP, an additional peak is seen on the histogram. In Fig. 4.2, pIRES2EGFP shows an additional peak, which is centered at ~10^{3} relative fluorescent units. The additional peak represents the fraction of cells that are expressing EGFP protein at detectable levels. It has been demonstrated that there is a linear relationship between the cell's FL1-H fluorescence and the amount of EGFP protein in the cell, despite complex mechanics of EGFP processing (Subramanian and Srienc, 1996). Thus, the average fluorescence of the transfected population will be a direct measure of the average level of EGFP expression in this population.

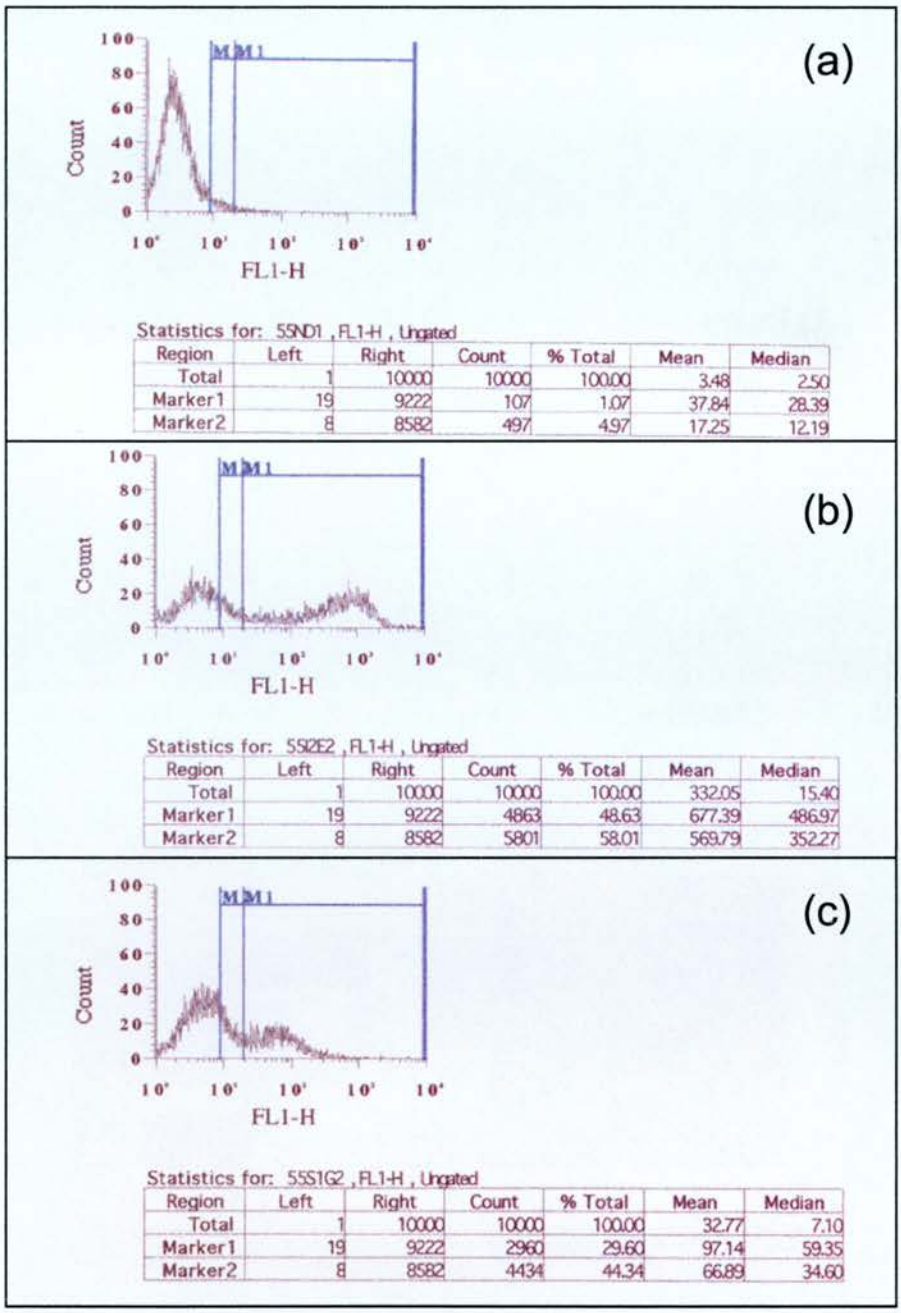


Figure 4.2. FACS histograms. COS-7 cells were either (a) left untransfected, or transfected with (b) pIRES2EGFPor (c) pS1G, complexed with Saint-Mix™. At 48 hours FACS was performed (see Methods). This figure shows FL1-H histograms of the FACS data. Note the logarithmic scale of FL1-H intensity. Regions M1 and M2 were created to encompass the approximate upper 5% and 1% of cells in the untransfected sample, respectively. These regions were then applied to the transfected samples, for statistical analysis.

Several factors influence the number of cells within this additional peak and their mean fluorescence. The promoter driving EGFP expression will influence the amount of protein expressed per cell and thus mean fluorescence. Protein fusions with EGFP may affect the EGFP chromophore, altering the amount of fluorescence generated per molecule of EGFP and hence the mean fluorescence. If expression is driven by an internal ribosome entry site, this will reduce the number of EGFP molecules translated per mRNA molecule, as initiation of translation from an IRES is less efficient than from the 5' end of the mRNA. In addition, the type of vector used will affect transfection, for example, large PAC plasmids are less efficiently transfected than small plasmids (this is discussed further in Section 5.2.6).

If such factors reduce the mean fluorescence below a certain point, the peak of transfected cells begins to merge with the background peaks. For example in Fig. 4.2, transfection of the pS1G vector into the COS7 cell line produces a second peak on the FL1-H FACS histogram, that merges with the background peak (the pS1G plasmid encodes a CFTR-EGFP fusion protein, driven by a P_{CMV} promoter, and was a gift from Barbara Stevenson, MMC, University of Edinburgh). Thus, EGFP fluorescence becomes masked by autofluorescence: this is the Achilles heel of FACS histogram analysis. As the CFTR promoter is very weak, and thus likely to produce peaks with a low FL1-H mean, a traditional FACS histogram analysis is not best suited for this study.

One possible way to dissect out the transfected population from merged peaks would be by subtraction analysis. If the peak profile of an untransfected sample were subtracted from the peak profile of a transfected sample, the peak of transfected cells should remain, despite a spectral overlap between the two peaks. However, this method is inappropriate for GCVs, due to the small number of cells that appear to be transfected with these PAC vectors. A small number of cells with an FL1-H fluorescence only slightly above background would not be a very convincing proof of EGFP expression. In order to demonstrate convincing EGFP expression, there must be a robust, qualitative difference between the untransfected and transfected populations.

The autofluorescence of untransfected cells is highly variable. However, the background fluorescence appears to be consistent across the spectrum. In analysing a dotplot of logarithmic FL1-H (green) vs. logarithmic FL2-H (orange) fluorescence (Fig. 4.3), it appears that in any untransfected cell the green and orange fluorescence are approximately equal: cells with low to high background fluorescence form a diagonal line across the plot. However when cells are transfected with a plasmid expressing EGFP, an additional population appears where FL1-H > FL2-H. This population appears to represent cells expressing EGFP. A similar type of dotplot analysis has been previously reported (Blaauw et al, 2000).

An RT-PCR experiment was performed to prove that the group of cells with FL1-H> FL2-H are a population of cells expressing EGFP, and that the FL1-H fluorescence of these cells is indicative of the amount of EGFP expressed. The COS-7 cell line was transfected with the vector pEGFP-N, using the LID method. At 48-hours, cells were mechanically sorted by a FACSCalibur machine. Three regions were delineated for gating based on an FL1-H: FL2-H dotplot (Fig. 4.3). **"Population 1"** included cells with FL1-H=FL2-H, **"Population 2"** included cells where FL1-H>FL2-H, with low FL1-H (green) fluorescence, and **"Population 3"** included cells with FL1-H>FL2-H, with high FL1-H (green) fluorescence. Cells were collected for each of these regions; in addition, an untransfected sample was run through the FACSCalibur and a sample of these cells with FL1-H=FL2-H was collected. At least 6,000 cells were collected for each sample (volumes were subsequently adjusted to correct for cell number).

The RNA was extracted from these four samples; DNase treatment and cDNA synthesis were performed (see Methods). RT-PCR was performed in duplicate (see Methods), for both EGFP (using the primers EGFPs and EGFPas – Appendix A) and GAPDH (using the primers GAPDHexon 8 and GAPDHint – Appendix A). –AMV-RT duplicate RNA samples (where the AMV-reverse transcriptase was omitted from the cDNA synthesis step), and water-only samples were included as negative controls, and samples including known quantities of the pEGFP-N vector (Clontech) were included as positive controls. The products were electrophoresed down 1% TBE gels and a Southern Blot was performed. The filters were probed with ^{32}P-ATP

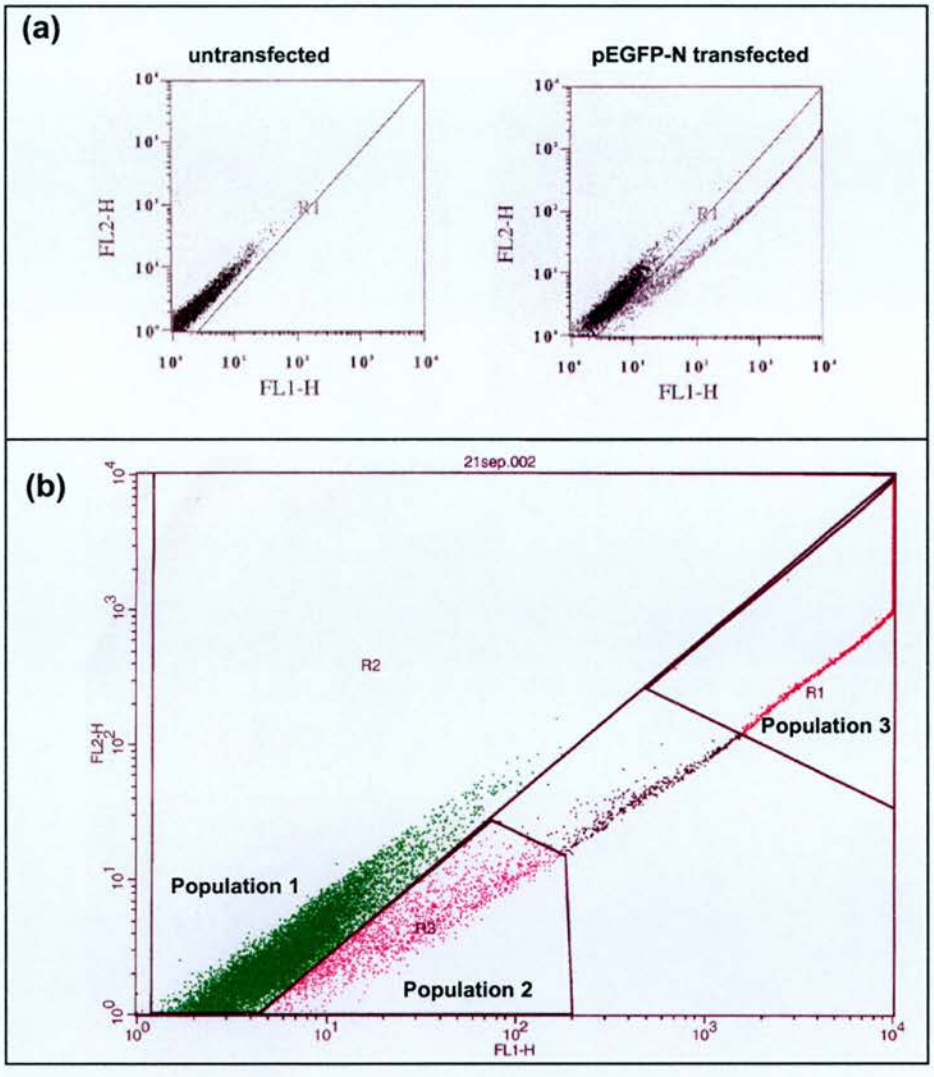


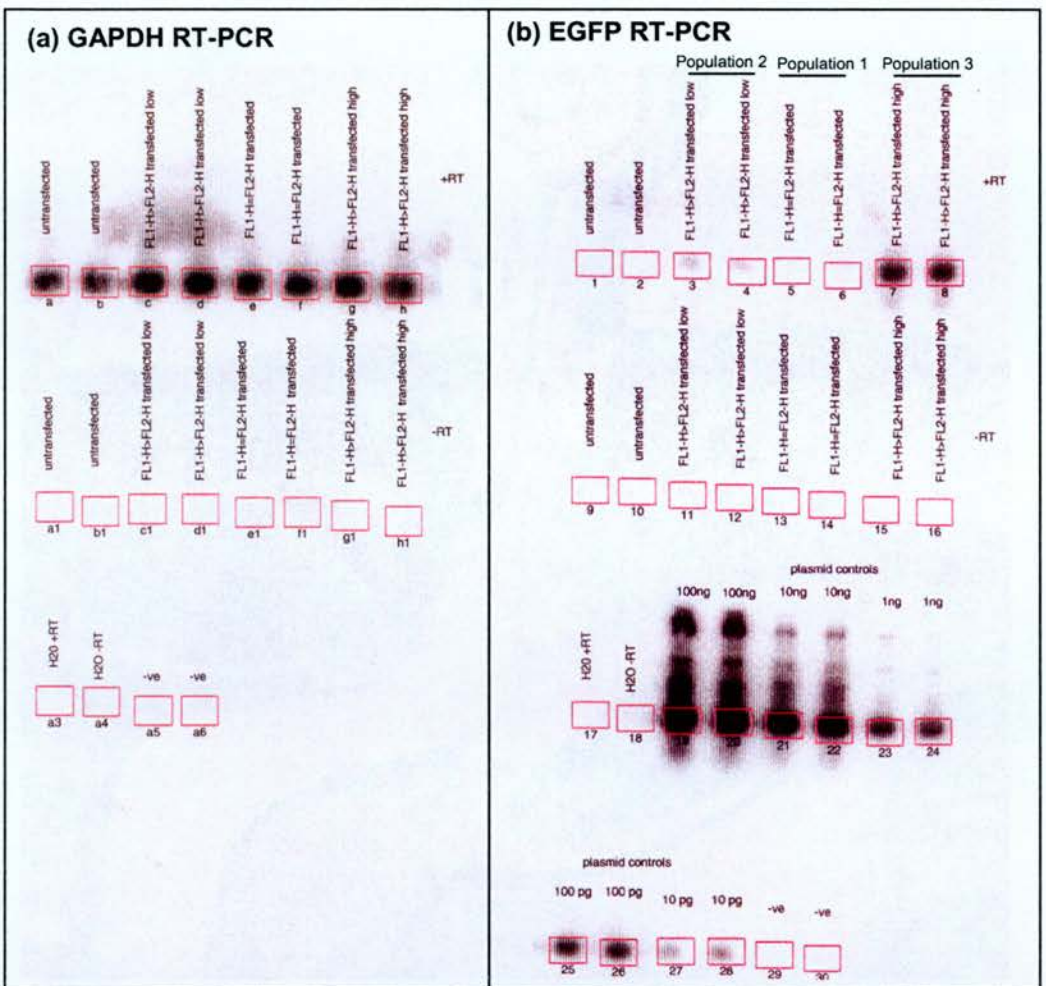
Figure 4.3. FACS FL1-H : FL2-H dotplot. (a) FACS dual dotplot analysis. COS-7 cells were transfected with the pEGFP-N vector, using the LID method, or left untransfected (as labelled). At 48-hours, FACS was performed; this figure shows the result of an FL1-H : FL2-H dotplot analysis. In the untransfected sample, all cells have an FL1-H reading approximately equal to their FL2-H reading. However, in the pEGFP-N transfected population, there is an additional population of cells, where FL1-H>FL2-H, these appear to be cells expressing EGFP. **(b) Gating populations of cells.** COS-7 cells were transfected with the pEGFP-N vector, using the LID method. At 48-hours, cells were prepared for FACS (see Methods). A FACSCalibur machine was used to analyse and mechanically sort the cells. Three regions were delineated for gating: **"Population 1"**, where FL1-H=FL2-H, **"Population 2"**, where FL1-H>FL2-H, with low fluorescence and **"Population 3"**, where FL1-H > FL2-H, with high fluorescence. In addition, an untransfected sample was run through the FACSCalibur machine, and an ungated sample was collected, to act as a control (not shown). At least 6,000 cells were collected for each sample.

labelled internal oligonucleotides (EGFPint for EGFP, and GAPDHint for the GAPDH – see Appendix A.)

Semi-quantitative analysis of ^{32}P signal was performed with a phosphoimager cassette (Fig. 4.4). GAPDH mRNA levels were similar between the four populations (it was not possible to perform a quantitative analysis of GAPDH levels, as no vector containing the GAPDH sequence was available in the lab). The phosphoimager values for the known quantities of pEGFP-N vector were plotted as a titration curve. These values were used to predict the number of EGFP cDNA copies in the three populations of gated cells, and in the untransfected sample (Fig. 4.4 and Fig. 4.5).

In addition, a T-test statistic was used to test whether the amount of cDNA in each population was significantly above the background levels of untransfected cells, and to compare the levels in "Population 2" and "Population 3" (see Appendix F1). In "Population 1", the predicted amount of EGFP cDNA was very low (1.45 amol). This was not significantly different from background levels. The predicted amount of EGFP cDNA was very high in "Population 3" (8372 amol.). A lower, but still substantial amount of EGFP cDNA was predicted in "Population 2" (16.9 amol), despite the fact that the FL1-H reading of cells in this region was comparable to that of untransfected cells. The amount of cDNA in "Population 3" and "Population 2" were both significantly different from the untransfected sample ($p = 0.0385$ and 0.0001, respectively). In addition, the amount of cDNA in "Population 3" was significantly different from that in "Population 2" ($p = 0.0023$). This shows that we can accurately separate cells expressing EGFP by their FL1-H: FL2-H ratio, and that there is a relationship between the FL1-H intensity of the population and the levels of EGFP mRNA present in that population.

There is a complex relationship between mRNA and protein levels, especially as proteins are generally more stable than their mRNAs (Carey and Smale, *Transcriptional Regulation in Eukaryotes*, Chapter 5). Hence, even if the phosphoimager results were a perfect measure of the mRNA levels, and the fluorescence was proportional to the amount of EGFP protein, we would not expect a



(c)

Control (amount plasmid)	Signal 1	Signal 2		Population	Signal	Predicted am't EGFP cDNA (amol.)
100 ng (32.2 fmol.)	764318.42	665096.34		Untransfected #1	786.09	0
10 ng (3.22 fmol.)	405934.20	371723.21		Untransfected #2	779.86	0
01 ng (0.322 fmol.)	132075.84	115855.00		Population 1 #1	2640.35	2.25
100 pg (0.0322 fmol.)	110630.95	111400.53		Population 1 #2	1839.26	0.64
-ve	1237.16	1028.12		Population 2 #1	10891.16	19.32
				Population 2 #2	8500.46	14.49
				Population 3 #1	145460.99	9016.00
				Population 3 #2	136471.42	7728.00

Figure 4.4. RT-PCR of FACSorted populations: phosphoimager results. RT-PCR was performed on the FACSorted populations shown in Fig. 4.2. The products were run down an agarose gel, and a Southern Blot was performed. The Filter was probed with ^{32}P-ATP labelled oligonucleotide (GAPDHint or EGFPint). A phosphoimager scan was used to detect radioactive signal. This top panel of this figure shows a picture of the phosphoimager scan of the filter. **(a) GAPDH RT-PCR:** the GAPDH signal looks similar in all of the +AMV-RT samples, and is absent from the -AMV-RT samples and H2O controls. **(b) EGFP RT-PCR**: In the +AMV-RT duplicates, a strong signal can be seen in the FLH-1>FLH-2 high samples ("population 3"), and in the plasmid controls. A weaker signal can be seen in the FLH-1>FLH-2 low samples ("population 2"). Little or no signal is seen in the FL1-H=FL2-H sample ("population 1") and the untransfected sample. No signal is detected in the -AMV-RT dulpicates or H2O controls. **(c)** Semi-quantitiative analysis was performed with imageQuant software. The predicted EGFP mRNA concentrations for each population were calculated based on a plot of the signal in the plasmid controls (as 1/20 of the total sample was used for each RT-PCR, this value was muliplied by 20 to calculate the predicted amount of cDNA in the original sample).

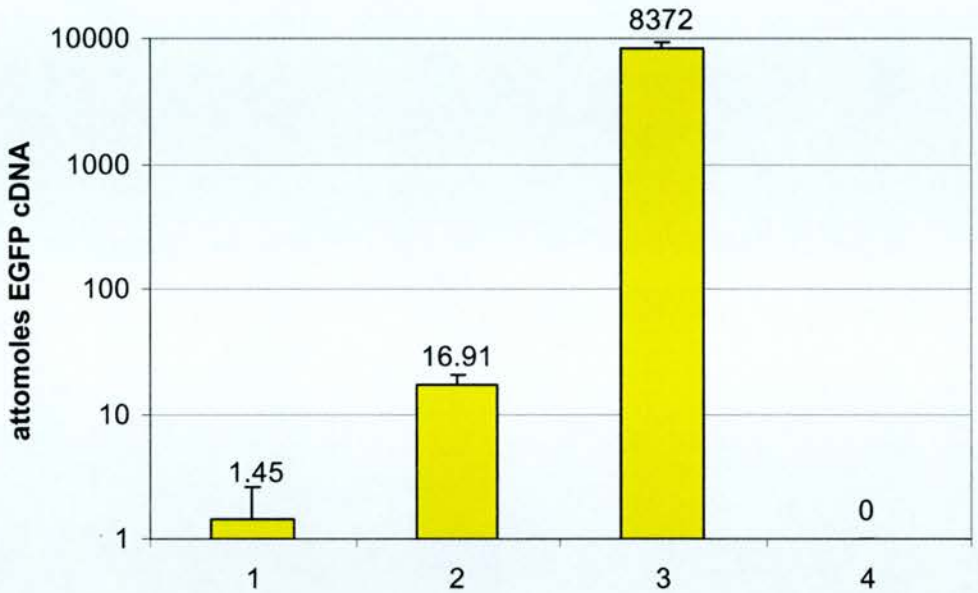


Figure 4.5. RT-PCR of FACSorted populations: Histogram of the phosphoimager results. Semi-quantitative RT-PCR was performed on the FACS sorted populations (Fig.4.2): a phosphoimager was used to calculate the predicted amount of EGFP cDNA in each population (Fig. 4.3). This figure shows a logarithmic histogram of the predicted cDNA concentrations in the three FACS sorted populations of the transfected sample: **(1)** "Population 1" (FL1-H =FL2-H) **(2)** "Population 2" (FL1-H>FL2-H with low FL1-H fluorescence) **(3)** "Population 3" (FL1-H>FL2-H with high FL1-H fluorescence) and **(4)** in the untransfected sample.

1:1 relationship between the mRNA levels and fluorescence intensity. However, based on these RT-PCR results, it seems likely that the FL1-H fluorescence intensity will continue to be a linear measure of EGFP protein levels (as shown by Subramanian and Srienc, 1996) when analysing FACS data by this dotplot method. It appears that even in cells where the FL1-H levels are not significantly above those of untransfected cells, the FL1-H: FL2-H ratio can be used to distinguish a population of cells expressing EGFP.

A mixing experiment was done to confirm that the number of cells within the delineated population (where FL1-H>FL2-H) is a consistent measure of the number of cells with detectable EGFP expression, across a range of values. The COS-7 cell line was transfected with the pEGFP-N plasmid, using the Saint-Mix™ reagent, or left untransfected. At 48 hours, the concentration of cells in each sample was quantified on a haemocytometer slide. The transfected sample was then mixed with the untransfected sample in different proportions (0 - all untransfected sample, 0.1, 0.25, 0.5, 0.75 and 1 – all transfected sample) and FACS analysis was performed on these samples. Fig. 4.6 shows a plot of dilution factor vs. the percentage of cells with detectable EGFP expression (the original dotplot data is shown in Appendix C). A curvilinear/linear relationship can be seen between the dilution factor and the percentage of cells with detectable EGFP expression, as measured by the dotplot, method. This simple experiment confirms that we can effectively measure the number of cells with detectable EGFP expression across a range of values, using this method.

The FL1-H: FL2-H dotplot method was used to analyse all subsequent FACS data. In each experiment, a dotplot of an untransfected sample was analysed first, in order to define a triangle excluding all cells in the untransfected sample (where FL1-H>FL2-H). This triangle region was then applied to the dotplots generated for each transfected sample. To quantify the percentage of cells with detectable EGFP expression, the proportion of cells within the triangle region was calculated. In order to quantify the average fluorescence of the transfected population, the mean fluorescence of cells within the triangle region was calculated. For untransfected samples, the mean of all cells (which fall outside the triangle region by definition)

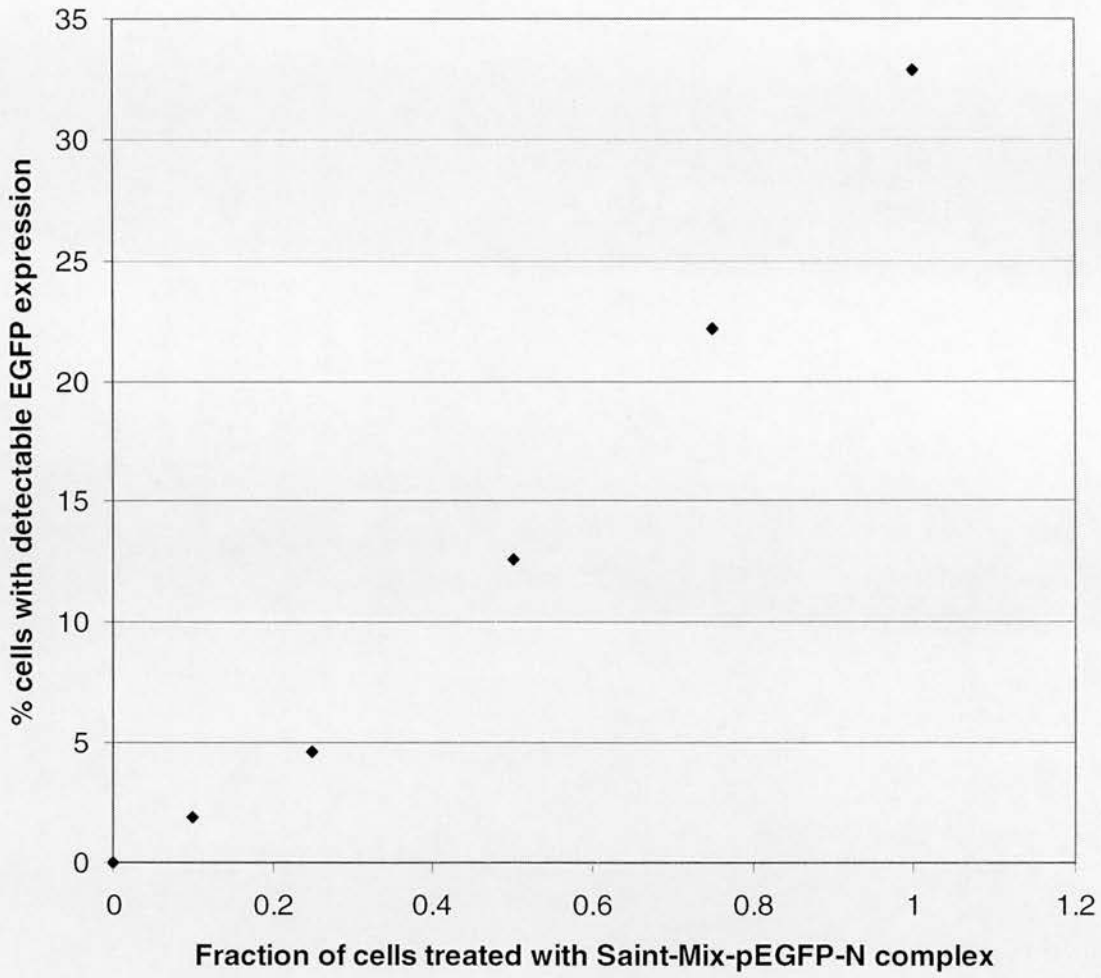


Figure 4.6. Dual-dotplot analysis: the effects of diluting the transfected sample. COS-7 cells were transfected with the pEGFP-N plasmid using the Saint-Mix™ reagent, or left untransfected. At 48-hours post-transfection, FACS was performed (see Methods), and the data was analysed by the dotplot method. The transfected sample was mixed with the untransfected sample in the following proportions: 0 (all untransfected), 0.1, 0.25, 0.5, 0.75, 1 (all transfected). A curvilinear/linear correlation appears to exist between fraction of cells that were transfected and the percentage of cells with detectable EGFP expression. The dotplots for this graph are shown in Appendix C.1.

was calculated as a measure of background fluorescence. For each experiment, duplicate or triplicate samples were processed, and the average values were calculated, regarding the percentage of cells with detectable EGFP expression (%[f]) and the average fluorescence of the transfected population (X_f). This data is displayed in the figures as histograms; the original dotplots are shown in Appendix C.

4.2.2 Saint-Mix™ vs. DOTAP

Saint-Mix™ and the liposome DOTAP were compared for transfection efficiency. Saint-Mix™ was complexed with the plasmid pCMVβ in the ratios (μl Saint-Mix™: μg plasmid) 20:1 10:1 or 5:1. DOTAP was complexed with the pCMVβ plasmid in the standard ratio 12 μl DOTAP: 1 μg plasmid (McLachlan et al, 1995). The complexes were prepared according to the standard protocol, then transfected into the COS-7 cell line (see Methods). 48 hours after transfection, the cells were lysed and a β-Galactosidase ELISA was performed to quantify β-Galactosidase protein in the cell lysates (see Methods). The results are shown in Figure 4.7. A 20:1 ratio of Saint-Mix™ to plasmid produced the optimum levels of transfection (352 ng/μl βgalactosidase), although similar results were obtained following Saint-Mix transfection at a ratio of 10:1 or 5:1 (267.23 and 206.63 ng/μl βgalactosidase, respectively), or DOTAP transfection (238.53 ng/μl βgalactosidase).

4.2.3 PEI vs. Saint-Mix™ vs. DOTAP

Polyethylenimine (PEI) transfection was performed according to the established protocol (Thorpe et al, 2002). COS-7 cells were transfected with the plasmid pCMVβ using either PEI, Saint-Mix™, or DOTAP (see Methods). 48 hours after transfection, cells were lysed and the βgalactosidase concentration in the cell lysate was measured with a luminometer assay (see Methods). The results are shown in Fig. 4.8. PEI produced the highest level of transfection (196.03 ng/μl βgalactosidase), which was significantly higher than Saint mix transfection (101.37 ng/μl) ($p=0.0076$), and DOTAP transfection (20.27 ng/μl)($p=0.0009$). In this experiment, Saint-Mix transfection was significantly higher than DOTAP transfection ($p=0.0095$).

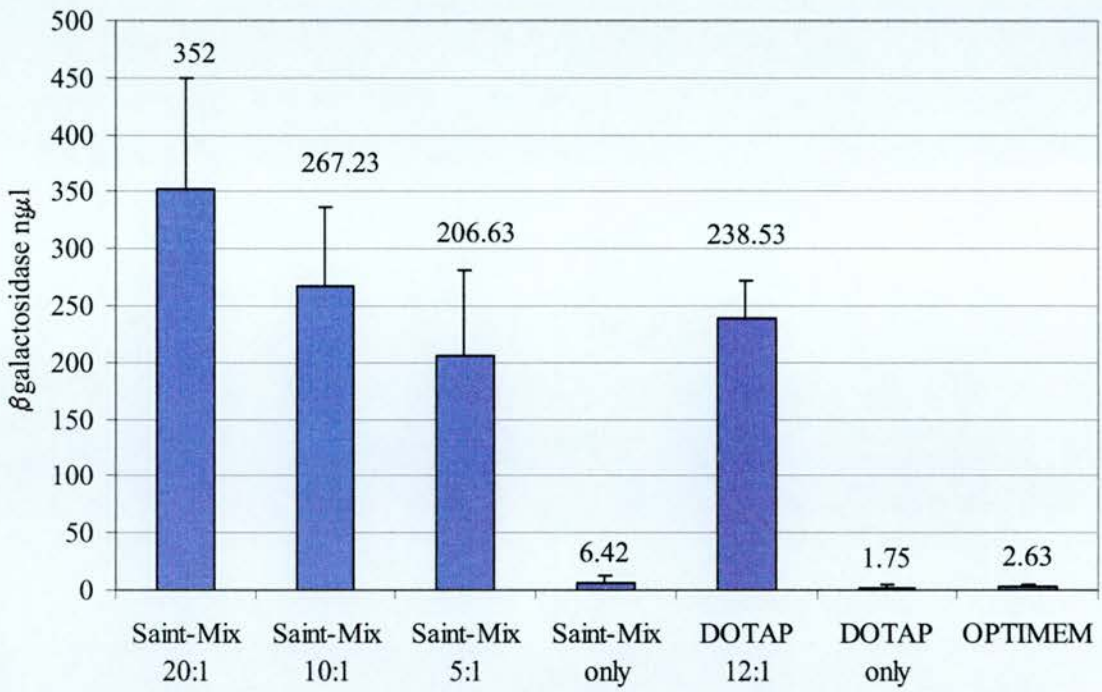


Figure 4.7. Saint-Mix™ ratio and DOTAP transfection of pCMVβ into the COS-7 cell line. The COS-7 cell line was transfected with the pCMVβ plasmid, complexed with Saint-Mix™ in a 20:1, 10:1 or 5:1 ratio (μl Saint-Mix™: μg DNA) or DOTAP in a 12:1 ratio (μl DOTAP: μg DNA)(as labelled). 48 hours after transfection, the βgalactosidase concentration was measured with a βgalactosidase ELISA (see Methods). This figure shows a histogram of the βgalactosidase concentration in the lysate.

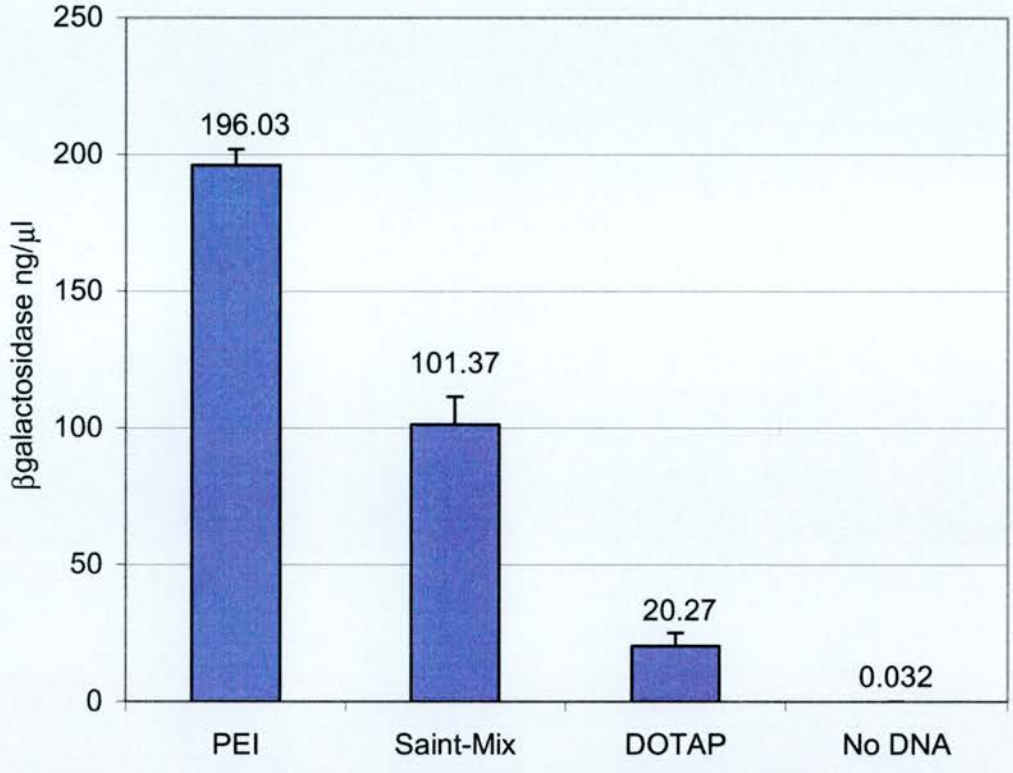


Figure 4.8. PEI, Saint-Mix™ and DOTAP transfection of pCMVβ into the COS-7 cell line. The COS-7 cell line was transfected with the pCMVβ plasmid, complexed with either PEI, Saint-Mix™ or DOTAP (as labelled). 48 hours after transfection, the βgalactosidase concentration was measured with a luminometer assay (see Methods). This figure shows a histogram of the βgalactosidase concentration in the lysate.

This experiment was repeated in the MDCK-IOWA cell line (Fig. 4.9); these results mirror those in the COS-7 cell line: PEI produced the optimum level of transfection (14.06 ng/µl βgalactosidase) which was significantly higher than Saint-Mix tranfection (3.96 ng/µl) (p=0.0097) and DOTAP transfection (0.91 ng/µl) (p=0.0058). Saint-Mix tranfection was significantly higher than DOTAP transfection (p=0.0005).

The experiment was repeated with EGFP-expressing plasmids, using FACS analysis, in order to dissect the number of transfected cells vs. the level of expression per cell. When COS-7 cells were transfected with the p1kbcfproEGFP plasmid (Fig. 4.10), Saint-Mix™ transfection produced the largest percentage of cells expressing EGFP (%[f]= 29.18). PEI and DOTAP were significantly less efficient (%[f] = 12.58 and 9.46, respectively)(p=0.0266 and 0.0157, respectively). However, the DOTAP, Saint-Mix™ and PEI reagents produced similar levels for the average fluorescence of the transfected population (X_f= 1153.08, 938.56 and 643.06, respectively). These results are at variance with those of the βgalactosidase assay.

When MDCK-IOWA cells were transfected with the pEGFP-N plasmid (Fig. 4.11), again Saint-Mix™ transfection produced the largest percentage of cells expressing EGFP (%[f] = 7.65), however, the number produced by PEI transfection (%[f] = 3.84) and DOTAP transfection (%[f] = 0.26) were not significantly lower. The average fluorescence of the transfected population was significantly higher with PEI (X_f = 2570.59) than with Saint-Mix™ (X_f = 391.82) (p=0.0262) and DOTAP (X_f =179.84)(p =0.0177).

Thus, Saint-Mix™ transfection appears to produce more cells expressing EGFP than PEI transfection, at least in the COS-7 cell line. However, following transfection of the MDCK-IOWA cell line with the pEGFP-N plasmid, the average fluorescence of the transfected population was much higher with PEI than with Saint-Mix™: this might provide a clue to explain the higher readings obtained from PEI transfection in the βgalactosidase assay (this is considered further in the Discussion).

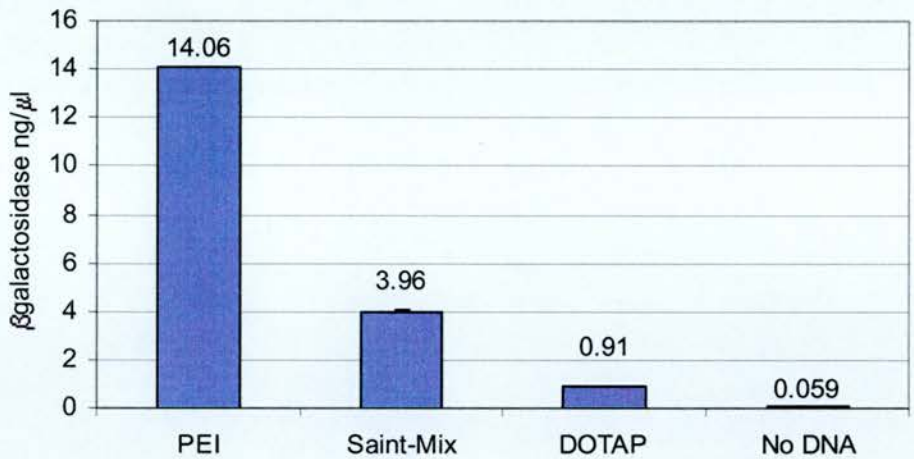


Figure 4.9. PEI, Saint-Mix and DOTAP transfection of pCMVβ into the MDCK-IOWA cell line. The MDCK-IOWA cell line was transfected with the pCMVβ plasmid, complexed with either PEI, Saint-Mix or DOTAP (as labelled). 48 hours after transfection, the βgalactosidase concentration was measured with a luminometer assay (see Methods). This figure shows a histogram of the βgalactosidase concentration in the lysate.

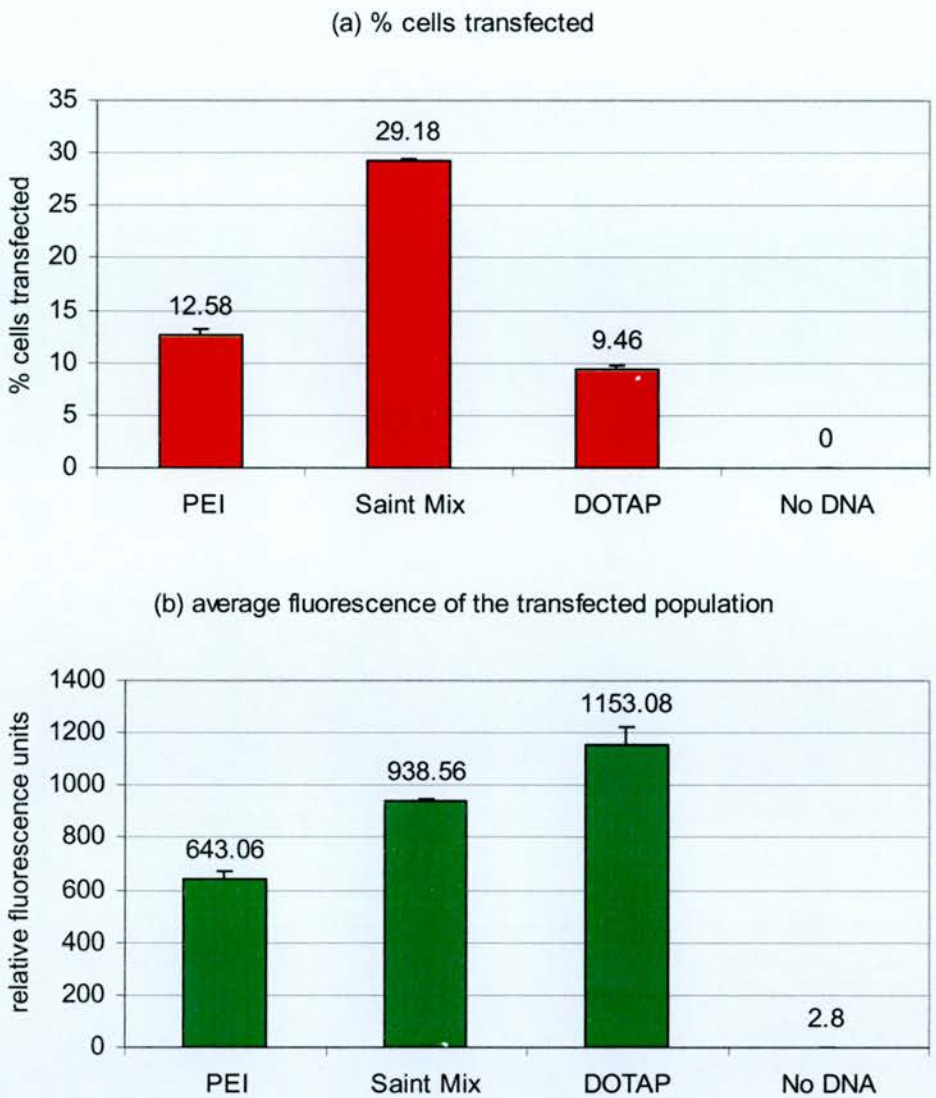


Figure 4.10. PEI, Saint-Mix and DOTAP transfection of p1kbcfproEGFP into the COS-7 cell line. The COS-7 cell line was transfected with the p1kbcfproEGFP plasmid complexed with either PEI, Saint-Mix or DOTAP (as labelled). At 48-hours post-transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.2.

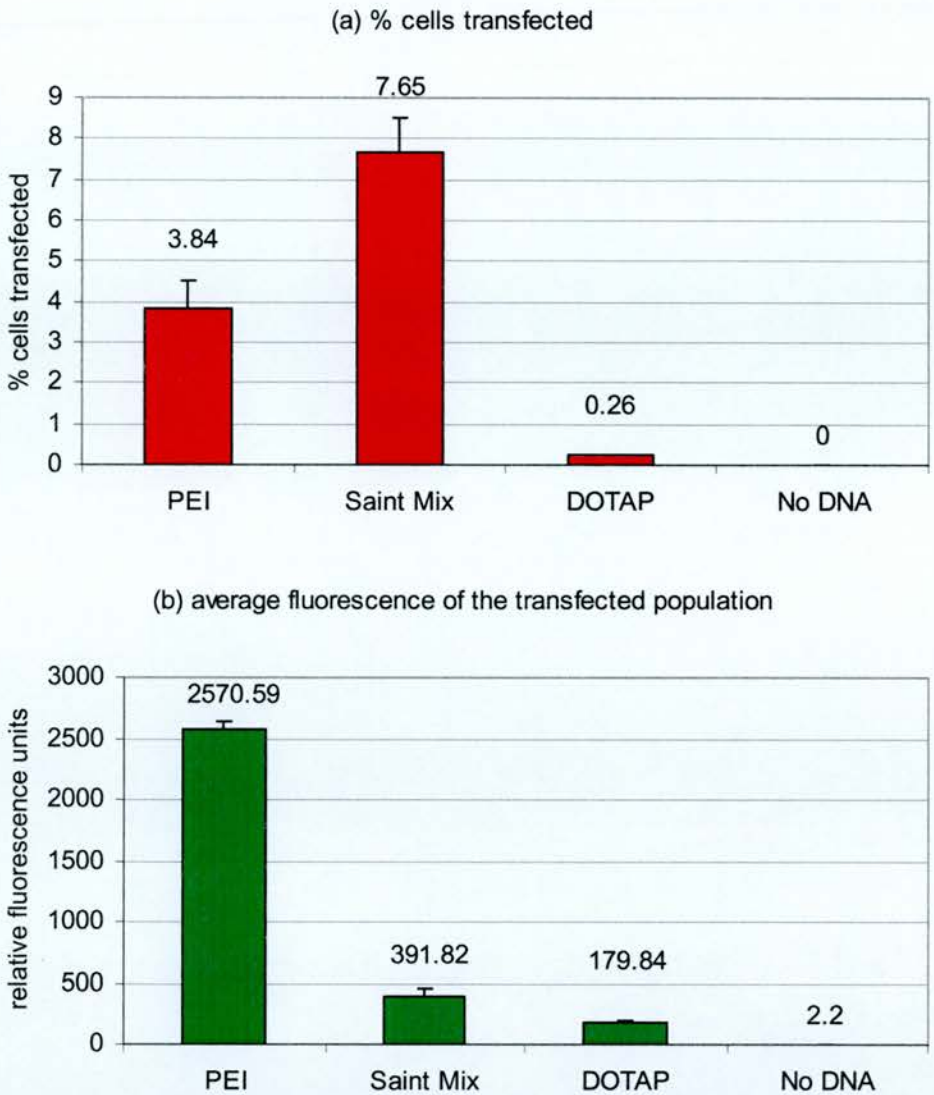


Figure 4.11. PEI, Saint-Mix and DOTAP transfection of pEGFP-N into the MDCK-IOWA cell line. The MDCK-IOWA cell line was transfected with the pEGFP-N plasmid complexed with either PEI, Saint-Mix or DOTAP (as labelled). At 48-hours post-transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.3.

4.2.4 The effects of P6: Lipofectin and Saint-Mix™ vs. LID and SID

Dr. Steve Hart, Institute of Child Health, London devised the LID method, which combines liposome, integrin-targeting peptide, and DNA for transfection (Hart et al, 1995 and 1998). Because epithelial cells of the lung express the integrin receptor that binds the P6 motif, this method has the potential to target this cell type in gene therapy. However, recent evidence suggests that the improved transfection efficiency with LID vector incorporating the P6 ligand may not in fact be accomplished by such a specific ligand-receptor interaction, as randomised P6 peptide sequences are equally efficacious (personal communication by Dr. Steve Hart).

LID complex was prepared according to the established method (personal communication from Dr. Steve Hart, see Methods). In addition, 20 µl of the Saint-Mix™ reagent was complexed with 1, 2, 4 or 8 µg P6 and 1 µg DNA; this was called the SID method. These two methods were compared to transfection with Saint-Mix™ or Lipofectin only. Initially, cells were transfected with pCMVβ, and a luminometer assay was used to analyse the amount of βgalactosidase in the cell lysate. These compounds were tested for transfection in the COS-7, MDCK-IOWA and Caco-2 cell lines.

In the COS-7 cell line (Fig. 4.12), the LID reagent produced the highest level of transfection (133.5 ng/µl βgalactosidase), which was significantly higher than SID with 2 µg P6 (31.92 ng/µl)(p=0.0199) and Lipofectin (1.55 ng/µl)(p=0.0115). Amongst the SID treatments, 2 µg P6 gave the highest level of transfection, although the levels obtained with 1 µg, 4 µg and 8 µg P6 and Saint-Mix only, were not significantly lower (13.55, 28.4, 24.7, and 10.55 ng/µl, respectively).

In the MDCK-IOWA cell line (Fig. 4.13), SID with 1 µg P6 produced the highest level of transfection (6.8 ng/µl βgalactosidase), although this value was not significantly different from that obtained with LID (4.25 ng/µl). SID with 1 µg P6 did exceed transfection with Lipofectin (1 ng/µl) (p=0.0386). Amongst the SID treatments, SID with 1 µg P6 was not significantly higher than SID with 2 µg P6, 4 µg P6, 8 µg P6 or Saint-Mix only (5.9, 6.55, 5.75 and 2 ng/µl, respectively).

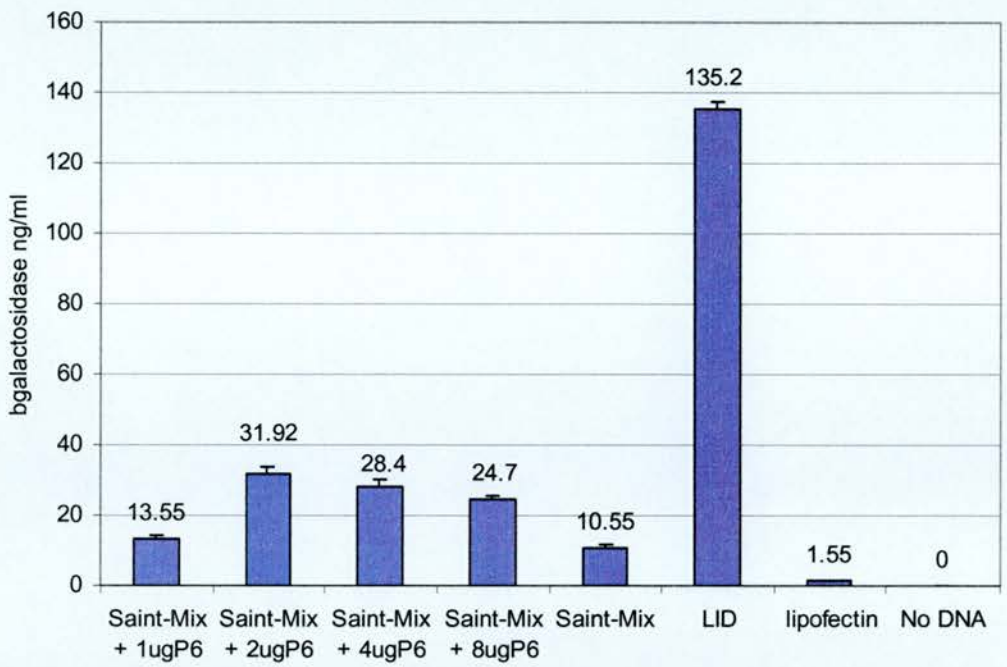


Figure 4.12. SID, Saint-Mix™, LID and Lipofectin transfection of pCMVβ into the COS-7 cell line. The COS-7 cell line was transfected with the pCMVβ plasmid, complexed with SID (Saint-Mix™ +1, 2, 4, or 8 μg P6), Saint-Mix™ only, LID or Lipofectin, or left untransfected (as labelled). 48 hours after transfection, the βgalactosidase concentration was measured with a luminometer assay (see Methods). This figure shows a histogram of the βgalactosidase concentration in the cell lysate.

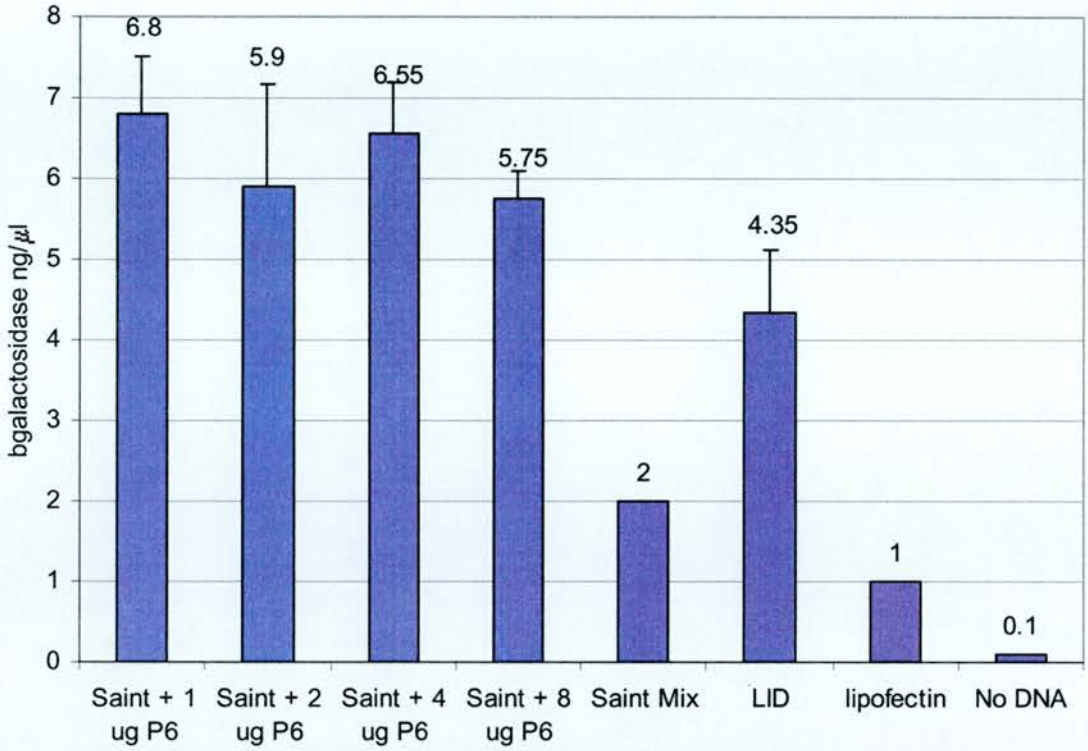


Figure 4.13. SID, Saint-Mix™, LID and Lipofectin transfection of pCMVβ into the MDCK-IOWA cell line. The MDCK-IOWA cell line was transfected with the pCMVβ plasmid, complexed with SID (Saint-Mix™ and 1, 2, 4, or 8 µg P6) Saint-Mix™ only, LID or Lipofectin, or left untransfected (as labelled). 48 hours after transfection, the βgalactosidase concentration was measured with a luminometer assay (see Methods). This figure shows a histogram of the βgalactosidase concentration in the cell lysate.

In the Caco-2 cell line (Fig. 4.14), SID with 1 μg P6 produced the optimum levels of tranfection, (5.53 ng/μl βgalactosidase), although the values obtained with LID and Saint-Mix only were not significantly different (4.95, and 1.45 ng/μl βgalactosidase, respectively). SID with 1 μg P6 transaction was, however, significantly more efficient than Lipofectin transfection (0.11 ng/μl) (p=0.0145). Amongst the SID treatments, SID with 1 μg P6 produced a significantly higher level of transfection than SID with 2 μg P6 or 4 μg P6 (2.23 and 2.24 ng/μl, respectively)(p=0.0481 and 00487, respectively).

Subsequently, cells were transfected with the pEGFP-N plasmid and FACS analysis was used to determine the number of cells expressing EGFP and the mean fluorescence of the transfected population. The SID method (using the optimum amount of P6 for each line, as determined by the previous β–galactosidase assay) was compared with the LID method, Saint-Mix™ alone, and Lipofectin alone.

In the COS-7 cell line (Fig. 4.15), LID, Saint-Mix and SID produced a similar percentage of cells expressing EGFP (%[f] = 28.88, 22.38, 20.07, respectively); each of these reagent produced a significantly higher percentage than Lipofectin (%[f] = 2.84)(p=0.0745, 0.0111 and 0.0011, respectively). However, Saint-Mix, LID, Lipofectin and SID produced a similar mean fluorescence amongst the transfected population (X_f = 4139.5, 4077.4, 3615.96 and 3550.16, respectively).

In the MDCK-IOWA line (Fig. 4.16), the LID, Saint-Mix and SID methods produced a similar percentage of cells expressing EGFP (%[f] = 6.87, 6.0, and 4.91, respectively), while Lipofectin produced a significantly lower percentage (%[f] 1.37) than LID (p=0.0082). The LID and SID methods produced a similar mean fluorescence amongst the transfected population (X_f = 1888.24 and 1590.75). Lower levels were found in the Saint Mix (X_f = 215.24)(compared with SID, p=0.0325) and lipofectin (X_f = 70.23)(compared with LID, p=0.001) transfections.

In Caco-2 (Fig. 4.17), LID transfection (%[f] = 8.03) produced a significantly higher percentage of cells expressing EGFP than SID (%[f] = 6.11)(p=0.0445), Saint-Mix (%[f] = 4.67)(p=0.033), and Lipofectin (%[f] =0.55)(p=0.0037). Furthermore, SID

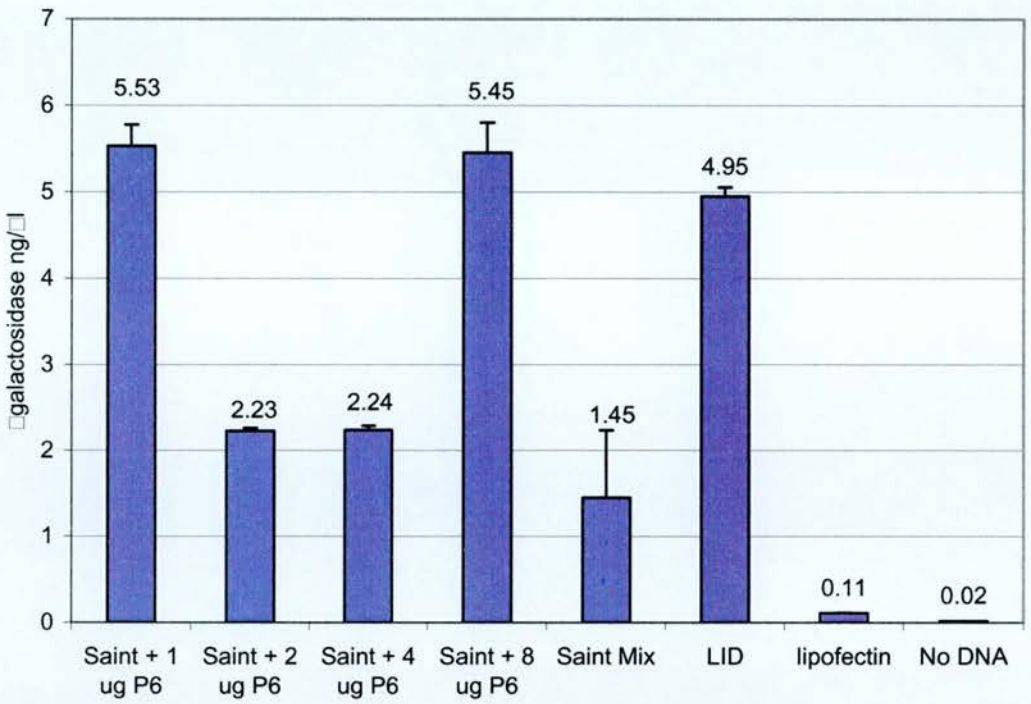


Figure 4.14. SID, Saint-Mix™, LID and Lipofectin transfection of pCMVβ into the Caco-2 cell line. The Caco-2 cell line was transfected with the pCMVβ plasmid, complexed with SID (Saint-Mix + 1, 2, 4, or 8 μg P6), Saint-Mix™ only, LID or Lipofectin, or left untransfected (as labelled). 48 hours after transfection, the βgalactosidase concentration was measured with a luminometer assay (see Methods). This figure shows a histogram of the βgalactosidase concentration in the cell lysate.

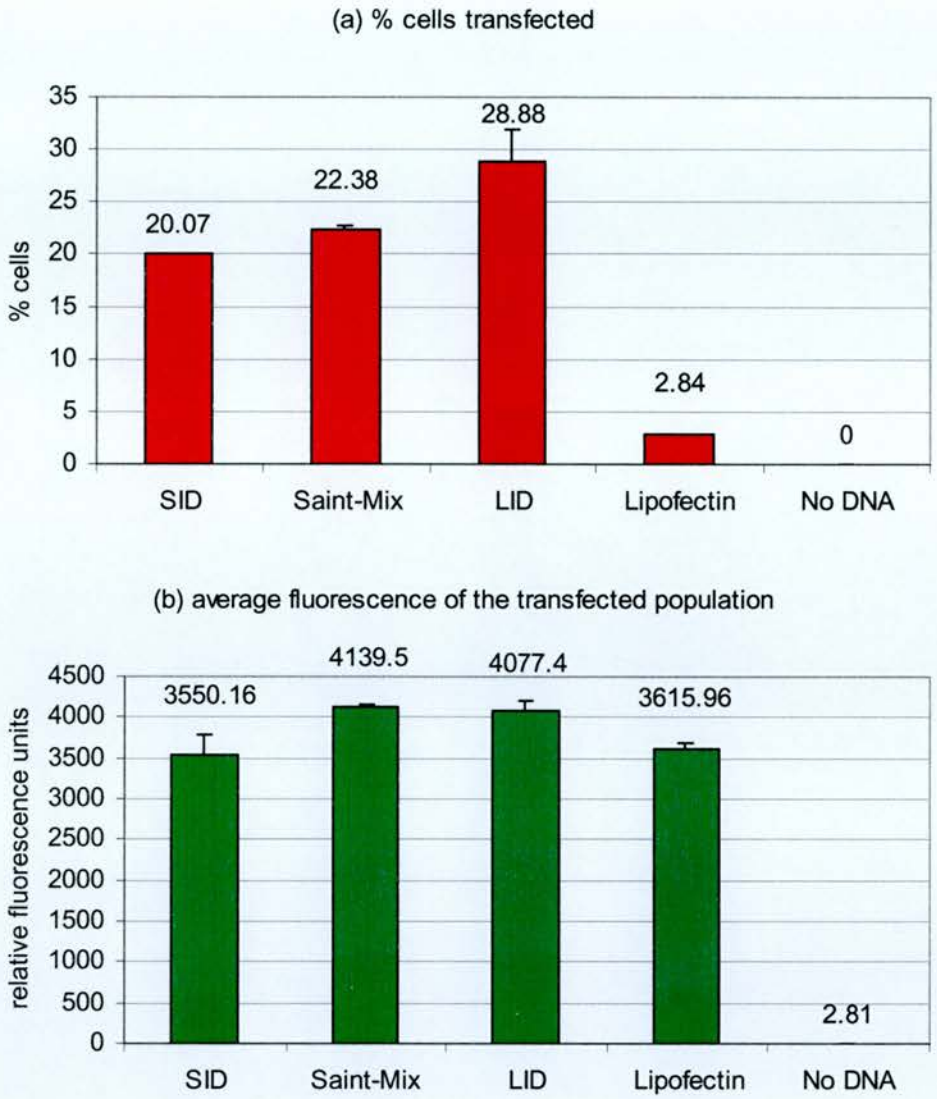


Figure 4.15. SID, Saint-Mix™, LID and Lipofectin transfection of pEGFP-N into the COS-7 cell line. The COS-7 cell line was transfected with the pEGFP-N plasmid complexed with either SID, Saint-Mix™, LID or Lipofectin (as labelled). 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.4.

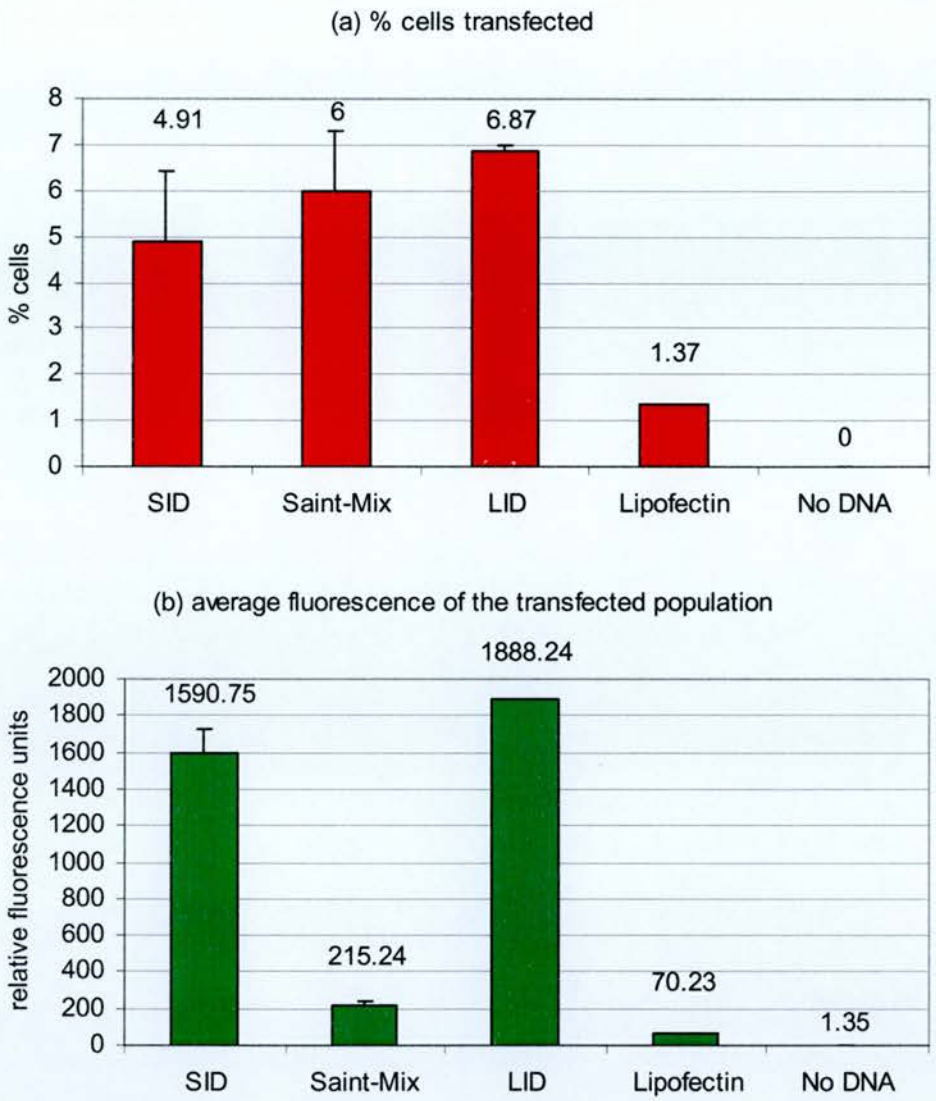


Figure 4.16. SID, Saint-Mix™, LID and lipofectin transfection of pEGFP-N into the MDCK-IOWA cell line. The MDCK-IOWA cell line was transfected with the pEGFP-N plasmid complexed with either SID, Saint-Mix™, LID or Lipofectin (as labelled). 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.5.

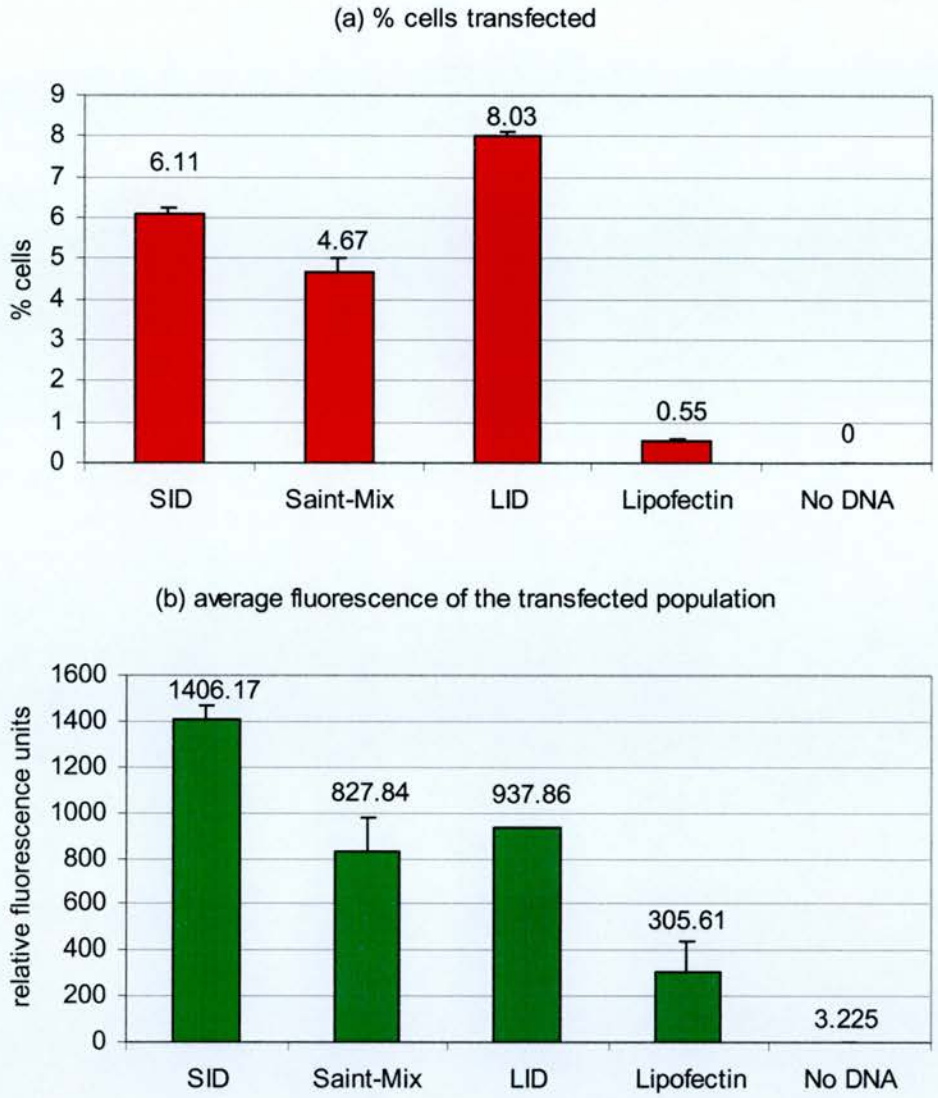


Figure 4.17. SID, Saint-Mix™, LID and Lipofectin transfection of pEGFP-N into the Caco-2 cell line. The Caco-2 cell line was transfected with the pEGFP-N plasmid complexed with either SID, Saint-Mix, LID or Lipofectin. At 48-hours post-transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.6.

and Saint-mix produced a significantly higher percentage than Lipofectin(p=0.0078 and 0.0262, respectively). However, SID produced the highest average fluorescence amongst the transfected population (X_f =1406.17), which was significantly higher than LID (X_f =937.86)(p=0.0441) and Lipofectin (X_f =305.61)(p=0.0413), but not significantly different from Saint-Mix (X_f =827.84).

Thus, the FACS results were not consistent with the βgalactosidase assay results. This was probably the result of a general decrease in the efficiency of the Saint-Mix™ reagent throughout the course of the project (see Discussion). It was evident that the total amount of protein generated in a transfected sample derived from a complex interaction between the number of cells expressing the protein and the level of expression per cell. It was also apparent that each cell line had a unique profile of transfection efficiency.

FACS was also used to compare these methods in the T84 cell line. The pEGFP-N plasmid was transfected into this cell line using either SID (with 1 μg P6), LID, Saint- Mix™ or DOTAP (Fig. 4.18). In this line, LID produced the highest percentage of cells expressing EGFP (%[f] = 21.56), which was significantly higher than SID(%[f] = 15.62)(p=0.0164), Saint-Mix (%[f] = 14.05)(p=0.0244) and DOTAP (%[f] = 8.53)(p=0.0021); furthermore, SID produced a significantly higher level than DOTAP (p = 0.0094). However, the average fluorescence of the transfected population was similar for Saint-Mix, SID, LID and DOTAP (X_f = 440.65, 373.16, 365.1 and 331.63, respectively).

Finally, transfection efficiency was optimised in the HBE cell line. SID, Saint-Mix™, LID, and DOTAP were used to transfect the pEGFP-N vector into this line (Fig. 4.19). Saint-Mix produced the highest percentage of cells expressing EGFP (%[f] = 13.23), which was simlar to SID and LID (%[f] = 11.5 and 7.73). SID produced a significantly higher level than DOTAP (%[f] = 2.69)(p=0.0366). The average fluorescence of the transfected population was similar between LID, Saint-Mix, DOTAP and SID (X_f = 1021.62, 896.9, 824.08 and 777.38, respectively).

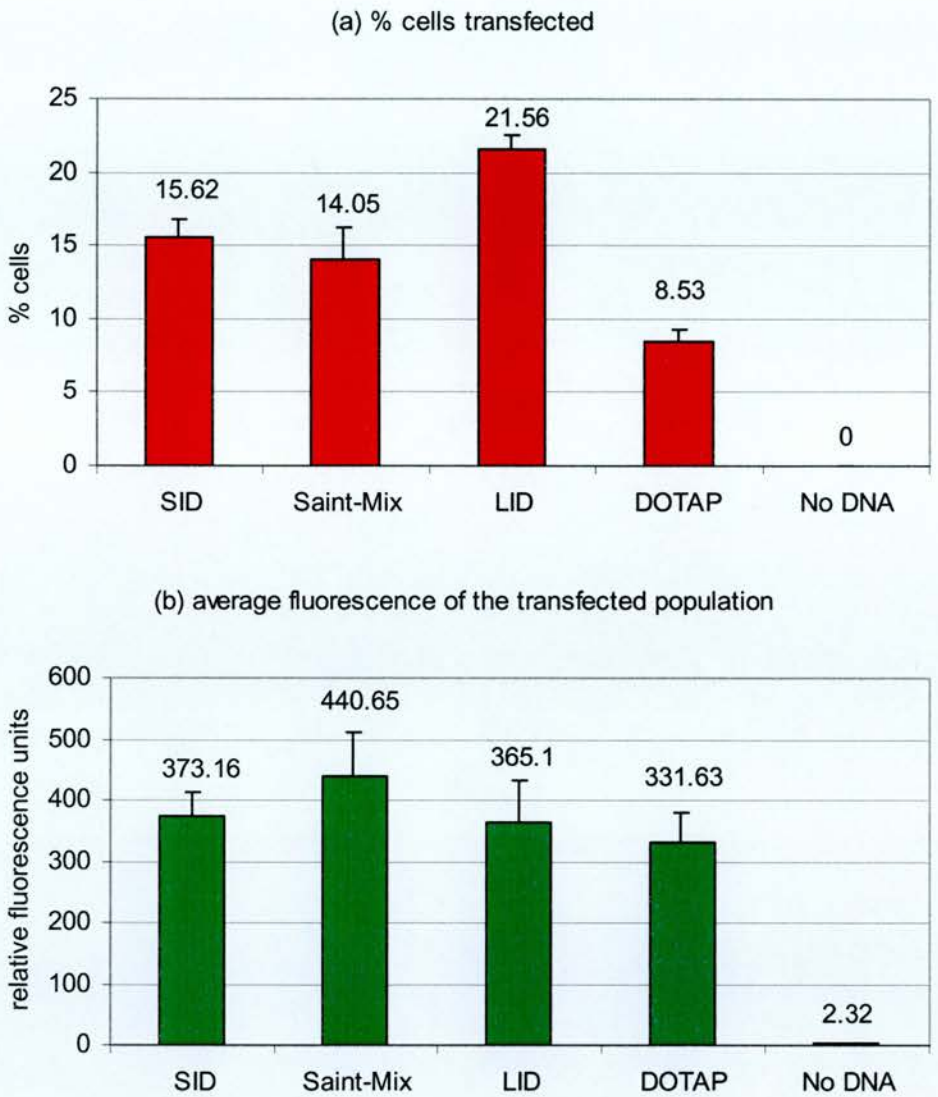


Figure 4.18. SID, Saint-Mix™, LID and DOTAP transfection of pEGFP-N into the T84 cell line. The T84 cell line was transfected with the pEGFP-N plasmid complexed with either SID, Saint-Mix™, LID or DOTAP (as labelled). 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.7.

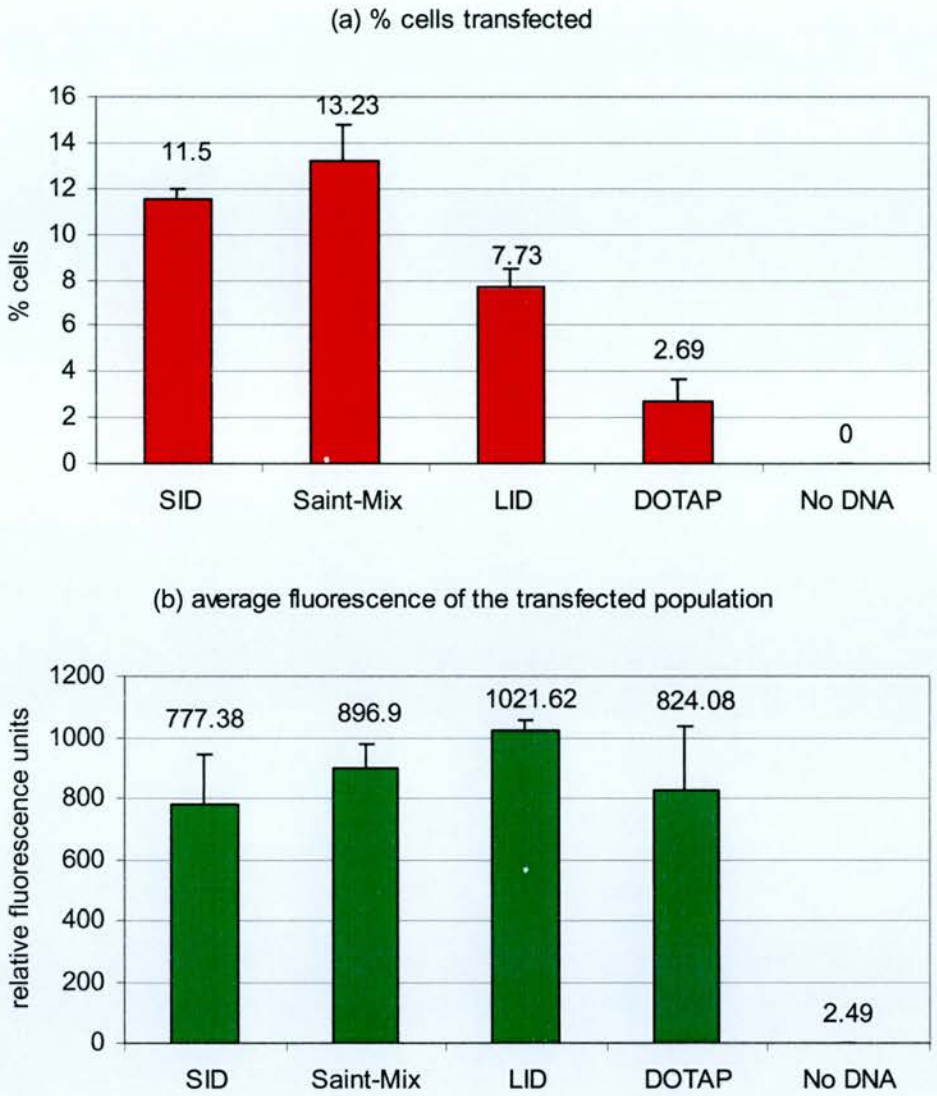


Figure 4.19. SID, Saint-Mix™, LID and DOTAP transfection of pEGFP-N into the HBE cell line. The HBE cell line was transfected with the pEGFP-N plasmid complexed with either SID, Saint-Mix™, LID or DOTAP (as labelled). 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.8.

4.3 Discussion

Experiments confirm that an FL1-H: FL2-H dotplot can be used to distinguish cells expressing EGFP in a transfected sample. An RT-PCR experiment confirms that the population of cells with FL1-H>FL2-H express EGFP, and that there is a positive correlation between the FL1-H intensity and the amount of EGFP mRNA in these cells. Mixing experiments show that the number of cells in the region where FL1-H>FL2-H is an accurate measure of the number of cells expressing EGFP across a range of values.

Transfection of cell lines is notoriously inconsistent. The overall efficiency of transfection varies greatly from day to day due to a combination of factors, many unknown. One must be careful not to over-interpret the results of these studies.

In order for a cell to express protein from a transgenic vector, DNA must be taken up into the cell, enter the nucleus and become transcribed. Different methods accomplish this in different ways. For example, a lipoplex achieves cell entry through the process of endocytosis. Addition of the integrin receptor targeting peptide allows this endocytosis to become receptor-mediated. In the PEI method, DNA is condensed and thus is afforded some protection from endosome degradation following cell entry. Despite much research on the subject, many facets of the transfection process remain a mystery.

In comparing PEI, Saint-Mix™, and DOTAP (Section 4.2.3), some discrepancies were found between the results of pCMVβ transfection (as measured by a luminescent assay), and EGFP transfection (as measured by FACS). One clue as to this reason for this discrepancy may come from the experiment on MDCK-IOWA cells transfected with the pEGFP-N plasmid using either PEI, Saint-Mix™, or DOTAP (Fig. 4.11). Although Saint-Mix™ transfected the largest number of cells, the PEI method produced a much higher mean fluorescence amongst the transfected population. This is likely to account for the apparent increase in transfection efficiency with PEI over that found with Saint-Mix™, as reported by the βgalactosidase luminometer assay.

This result was not mirrored when the p1kbcfproEGFP vector was transfected into the COS-7 cell line (Fig. 4.10). Thus, the effect may be specific to the P_{CMV} promoter. It is possible that a tightly regulated promoter, such as that of CFTR, would not be amenable to indiscriminate upregulation of expression, whilst the promoter of a virus, which has evolved to allow maximum expression levels, would be amenable to such upregulation. Further study would be needed to evaluate this hypothesis. In this instance FACS analysis of EGFP expression proved to be superior to βgalactosidase protein assays, as it provided a vital insight into the components of transfection efficiency.

In comparing transfection reagents it became obvious that the overall transfection efficiency (e.g. amount of protein expressed) is made up of a complex interaction between the percentage of cells expressing the transgene and the level of expression per cell. These components of transfection appear to be affected in different ways by different transfection reagents.

The ultimate aim of comparing transfection reagents in this study was to maximise the number of cells transfected, in order to facilitate a study of CFTR control elements, and their effect on transgene expression. Although PEI may increase the level of expression in certain situations. this effect may not be consistent between promoters. This effect may be due to a change in the level of expression per transfected cell, rather than a change in the number of cells expressing the transgene. The aim of the GCV approach is to attempt to replicate the endogenous CFTR expression levels, not to increase them, thus PEI may not be an appropriate tranfection reagent for these studies.

Although the Saint-Mix™ method showed promise at the start of the project, this reagent produced variable results, and the efficiency of transfection with this reagent generally declined through the course of the project. For example, at the start of the project it was possible to transfect ~50% of cells in the COS-7 cell line with this reagent (Fig. 4.2), but by the end of the project only ~20-30% of cells were transfected in the COS-7 cell line with this reagent (Fig. 4.15). In addition, a few inconsistent batches of Saint-Mix™ were obtained, which produced little or no

transfection at all. It is unclear whether these fluctuations were due to changes in product quality (for example due to batch to batch variation, or changes in the product formulation or shipping conditions), or whether they were an artifact of the system (for example, this transfection method might be particularly sensitive to changes in composition between batches of culture media or serum, or the CO_2 level in the incubator).

When Saint-Mix™ worked at its best, it produced a very high level of transfection, and this was primarily due to an increase in the number of cells expressing the transgene. If the causes of variation in transfection efficiency could be pinpointed and controlled, Saint-Mix™ would be an excellent transfection reagent for this type of study. The integrin-targeting peptide P6 may have the potential to increase transfection efficiency even further in concert with Saint-Mix™, in the SID method. However, the SID method showed the same variability as Saint-Mix™ alone. The overall decrease in Saint-Mix™ reagent efficiency probably explains the discrepancy between the βgalactosidase and FACS results, when comparing the SID, Saint-Mix™, LID and Lipofectin reagents. Whatever the cause for such variations, they are a major barrier to using Saint-Mix™ for this project. Experiments to study the effects of CFTR promoter elements (and ultimately gene therapy formulations) will require a robust transfection system that produces consistent results.

The LID transfection method produced consistent results and transfected the greatest percentage of cells in every line analysed, except HBE. It is particularly attractive to use a method incorporating the P6 peptide, as this may generate some targeting specificity for epithelial cells of the lung (Hart, 1999), although the equivalent results with randomised P6 peptide call into question the mechanism by which P6 enhances transfection (see Section 4.2.4). Another advantage of the LID method is that the components are relatively inexpensive. In future gene therapy formulations, this could make the difference between an affordable therapy and a prohibitively expensive one.

To sum up, several reagents were analysed in order to optimise transfection efficiency. It was apparent that each method produced a unique profile amongst the

transfected population and that each method had its own strengths and weaknesses relating to the number of cells transfected, the level of expression per transfected cell, and consistency between experiments. The LID transfection system seems to be best suited to the purposes of this study, and was selected for further experiments. The FLH-1: FLH-2 dotplot method proved to be a superior way to analyse FACS data.

Chapter 5:Transfection of immortalised cell lines

5.1 Introduction

In order to analyse the effects of CFTR genomic context elements upon expression, the six vectors containing portions of the CFTR 5' region and/or intron 1, linked in a transcriptional fusion to the EGFP reporter gene (see Chapter 3) were transfected into permanent cell lines. The reader is advised to make use of the foldout vector diagram in Appendix E, for easy reference while reading this chapter.

Several cell lines were selected for expression analysis. The project commenced with the COS-7 cell line, due to its relative ease of transfection. This line was derived from the CV-1 cell line, isolated from African green monkey kidney (COS-7 was created by transforming the CV-1 line with an SV-40 origin-defective mutant)(Gluzman, 1991; ATCC website). This line exhibits fibroblast-like morphology. The COS-7 line does not express detectable levels of CFTR, and therefore one might expect expression from CFTR genomic context vectors to be restricted in this line.

Subsequently three cell lines that normally express CFTR were analysed. S.H. Madin and N.B. Darby derived the MDCK cell line from canine kidney cells in 1958 (Gaush et al, 1966; ATCC website). This line has an epithelial morphology and the MDCK-IOWA variant (a kind gift from David Sheppard, University of Bristol) expresses CFTR at high levels, thus one might expect permissive expression from CFTR genomic context vectors in this line.

The T84 cell line was created from a human colon tumour metastasis isolated from the lung (Dharmsathaphorn et al, 1984; ATCC website). This line expresses CFTR at high levels and has been used to study regulation of CFTR by cAMP agonists and PMA (Bargon et al, 1998). One might expect CFTR genomic context vectors to exhibit higher levels of expression in this line; particularly those containing the intron 1 DHS, which upregulates CFTR expression in gut cells (Smith, AN et al, 1996; Rowntree et al, 2001).

The Caco-2 cell line was isolated from a primary human colonic tumour, and demonstrates epithelial morphology (Fogh et al, 1977a; Fogh et al, 1977b; ATCC

website). As in the T84 line, one would expect upregulation of expression in response to the CFTR intron 1 DHS element, as these cells are derived from the gut.

Permanent epithelial cell lines derived from the lung are available, but these lines are comparatively difficult to propagate and transfect. Because of these difficulties, it was only possible to perform a preliminary study on one such line. The HBE line was derived from the human bronchus and exhibits epithelial morphology (Viallet et al, 1994).

The p1kbcfproEGFP vector was used to investigate the effects of CFTR proximal 5' elements upon expression. pEGFP-N is of similar structure, but the P_{CMV} promoter drives EGFP expression: this vector was used as a positive control.

In order to analyse the effects of additional, distal CFTR 5' elements upon expression, the $_{PAC}$65kbcfproEGFP vector (81 kb), containing 65 kb of CFTR 5' region, was compared to the p1kbcfproEGFP plasmid (4.9 kb). However, it would not be prudent to directly compare expression from vectors that differ so greatly in size and form. Preliminary experiments showed that large PACs exhibit reduced expression in comparison to small plasmid vectors (results not shown). Expression from the vectors $_{PAC}$RC2cmvEGFP and pEGFP-N was compared, in order to characterise and quantify the differences in transfection efficiency between small plasmids and large PACs. These vectors both contain the P_{CMV}–EGFP cassette, but $_{PAC}$RC2cmvEGFP is a 111 Kb PAC whereas pEGFP-N is a 4.7 kb plasmid.

The twin vectors $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP differ only in the absence or presence of CFTR intron 1, respectively. Expression from these vectors was compared, to determine the influence of the intron 1 DHS upon expression. EGFP translation is initiated from an internal ribosome entry site (IRES) in both vectors. The small plasmid pCFTRiresEGFP was included in the study as a control, to investigate the effects of initiating translation from an IRES, rather than the 5' end of the mRNA.

The LID method was selected for transfection of these permanent cell lines, for the reasons described in Chapter 4. High quality, uncontaminated plasmid DNA is

required for transfection. For these experiments, vector DNA was prepared by Qiagen maxiprep. Plasmid-safe DNase was used to digest any DNA from bacterial chromosomes that might be contaminating the preparation. Phenol-chloroform was used to extract plasmid DNA from the Plasmid-safe DNase reaction, and then a chloroform extraction was performed to remove residual phenol. Finally, the plasmid was ethanol precipitated and resuspended in sterile TE, pH8.0 (see Methods). When possible, the vectors used in an experiment were prepared on the same day, to minimise any batch-to-batch variations in transfection efficiency.

48 hours after the cells had been transfected, fluorescence activated cell sorting (FACS) was performed, and an FL1-H: FL2-H dotplot was created to dissect the percentage of cells expressing EGFP (%[f]) and the average FL1-H fluorescence of the transfected population (X_f), which is a measure of the level of EGFP expression (see Chapter 4). Triplicate samples were included in every experiment, and each experiment was performed on at least two, and usually three separate occasions, to confirm the results. Although the absolute levels of transfection varied between experiments, the trends in relative transfection efficiencies were consistent between experiments. For most samples, 10,000 cells were analysed by FACS. However, when the transfection efficiency was very low, a larger sample was analysed to increase the significance of the calculations. The original dotplots can be found in Appendix C (these also give details of the numbers of cells analysed by FACS for each sample).

An initial experiment comparing expression from pEGFP-N and $_{PAC}$RC2cmvEGFP demonstrated that large PAC vectors are less efficiently transfected than small plasmid vectors. Further experiments were done to investigate whether this effect was due to a difference in the vector copy number, and whether this effect could act in trans on a co-transfected small plasmid. These studies unveiled an interesting interaction: diluting a reporter plasmid with an anonymous plasmid did not necessarily reduce its transfection efficiency. Further experiments were done to investigate the mechanism of this phenomenon.

5.2 Results

5.2.1 Transfection of the COS-7 cell line

The COS-7 cell line was selected for the first transfection experiments as this line is easily transfected. COS-7 does not express CFTR at detectable levels, and therefore will act as an important control, to analyse whether genomic context elements down-regulate expression in an inappropriate cell type.

A panel of vectors was transfected into the COS-7 cell line (Fig. 5.1). Transfection of both the p1kbcfproEGFP and pEGFP-N vectors produced a large percentage of cells expressing EGFP (%[f] =37.39 and 33.7, respectively). This presumably reflects the relative ease of transfecting the COS-7 cell line with these small vectors. However, the average fluorescence of the transfected population was much lower with p1kbcfproEGFP (X_f =1090.52) than with pEGFP-N (X_f =4247.85). This is consistent with previous studies, which show that the CFTR proximal 5' region exhibits weak housekeeping-type promoter activity (Yoshimura et al, 1991b).

$_{PAC}$65kbcfproEGFP-transfected cells showed a substantial reduction in both the percentage of cells expressing EGFP (%[f] =1.86) and the average fluorescence intensity of the transfected population (X_f =110.08), in comparison with p1kbcfproEGFP-transfected cells. Two hypotheses were proposed to explain this reduction in expression: (1) the reduction might have been due to negative control elements in the region from –1kb to –65kb, (2) the reduction might have been an artifact of the large $_{PAC}$65kbcfproEGFP vector size. $_{PAC}$RC2cmvEGFP and pEGFP-N transfections were compared, to resolve this issue.

$_{PAC}$RC2cmvEGFP-transfected cells showed a reduction in both the percentage of cells expressing EGFP (%[f] =3.48) and the average fluorescence intensity of the transfected population (X_f = 552.62), in comparison with pEGFP-N-transfected cells. This suggests that a reduction in expression is intrinsic to large PAC vectors. As this reduction was comparable to that seen between the p1kbcfproEGFP and $_{PAC}$65kbcfproEGFP vector, there is no evidence that the additional region from -1 kb to –65 kb of CFTR 5' region restricts expression in the COS-7 cell line (see Discussion).

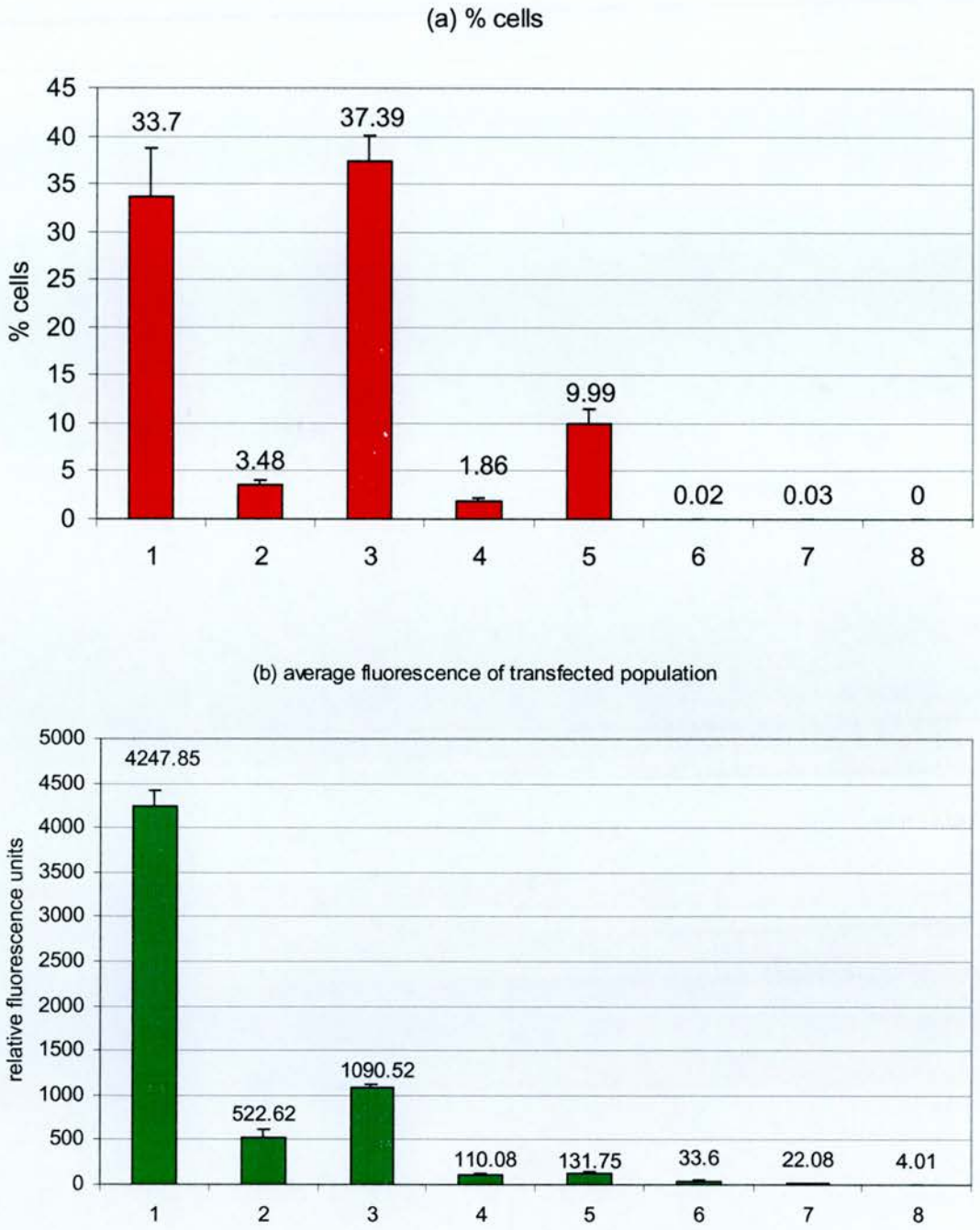


Figure 5.1. Transfection of EGFP reporter constructs into the COS-7 cell line. The COS-7 cell line was transfected with the plasmids **(1)** pEGFP-N, **(2)** $_{PAC}$RC2cmvEGFP, **(3)** p1kbcfproEGFP, **(4)** $_{PAC}$65kbcfproEGFP, **(5)** pCFTRiresEGFP, **(6)** $_{PAC}$RC1iresEGFP, and **(7)** $_{PAC}$RC2iresEGFP using the LID reagent, or **(8)** left untransfected. 48 after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.9.

In the COS-7 cell line, the vector size had a major effect upon both the percentage of cells expressing EGFP and the average fluorescence of the transfected population. The promoter present in the vector (e.g. CFTR 5' region of P_{CMV}) had a comparatively minor effect upon these parameters.

Next, the $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP transfections were compared, to assess the effects of the intron 1 DHS in the COS-7 line. No appreciable difference was seen between $_{PAC}$RC1iresEGFP (%[f] = 0.02, X_f = 33.6) and $_{PAC}$RC2iresEGFP (%[f] = 0.03, X_f = 22.08). Thus, the intron 1 element does not seem to alter expression in this line. This is consistent with a previous report, which shows that DNase hypersensitivity at the intron 1 site is lacking in cell lines that do not express CFTR, and this element does not appear to upregulate CFTR expression in such lines (Smith, AN et al, 1996).

pCFTRiresEGFP transfections showed a marked reduction in both the percentage of cells expressing EGFP (%[f] = 9.99) and the average fluorescence intensity of the transfected population (X_f = 131.75), in comparison to pEGFP-N. This suggests that initiation of translation from an IRES is much less efficient than from the 5' end of the mRNA molecule. Inefficient translation from an IRES, coupled with the effects of large vector size and a weak CFTR promoter may explain why the $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP transfection efficiencies were very low.

5.2.2 Transfection of the MDCK-IOWA cell line

Subsequently, the MDCK-IOWA cell line was transfected with these genomic context vectors (Fig. 5.2). Although this line is not of primary interest to CF gene therapy (as the kidney does not show disease phenotype in CF), this line expresses CFTR at high levels and is moderately easy to transfect. Thus, MDCK-IOWA cells may prove useful in dissecting CFTR regulatory elements.

In this line, p1kbcfproEGFP-transfected cells showed a moderate level of EGFP expression. However, both the percentage of cells expressing EGFP and the average fluorescence of the transfected population were reduced with p1kbcfproEGFP (%[f] = 6.92, X_f =388.46), in comparison to pEGFP-N (%[f]= 11.69, X_f =2022.72). The CFTR 5' region is acting as a weak promoter.

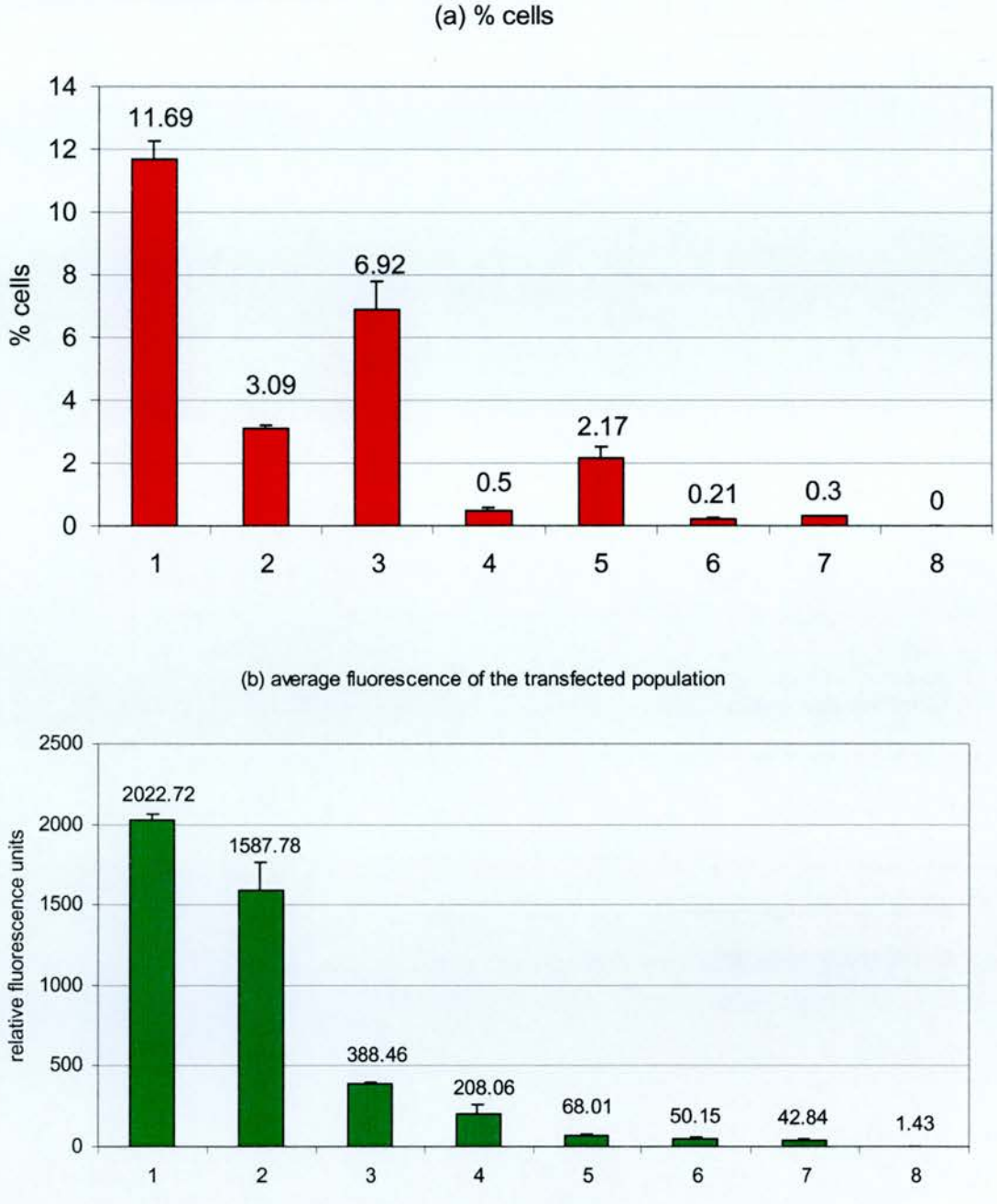


Figure 5.2. Transfection of EGFP reporter constructs into the MDCK-IOWA cell line. The MDCK-IOWA cell line was transfected with the plasmids **(1)** pEGFP-N, **(2)** $_{PAC}$RC2cmvEGFP, **(3)** p1kbcfproEGFP, **(4)** $_{PAC}$65kbcfproEGFP, **(5)** pCFTRiresEGFP, **(6)** $_{PAC}$RC1iresEGFP, and **(7)** $_{PAC}$RC2iresEGFP using the LID reagent, or **(8)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.10.

Again, $_{PAC}$65kbcfproEGFP-transfected cells showed reduced expression (%[f]=0.5, X_f =208.06) in comparison to p1kbcfproEGFP-transfected cells. However, this effect was mirrored when comparing $_{PAC}$RC2cmvEGFP transfections (%[f]=3.09, X_f =1587.78) to pEGFP-N transfections, and can thus be explained as an artifact of the large PAC vector context.

As in the COS-7 cell line, the vector size had a major effect upon the percentage of cells expressing EGFP, while the promoter type had a lesser effect. However, in contrast to COS-7, the vector size had only a minor effect upon the average fluorescence of the transfected population, while the promoter had a larger effect. This suggests that the promoter is playing a major role in directing the level of expression in the MDCK-IOWA cell line. Again, pCFTRiresEGFP-transfected cells showed a reduced level of expression (%[f] = 2.17, X_f = 68.01).

$_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP showed similar expression profiles following transfection (%[f]=0.21 and 0.3, X_f = 50.15 and 42.84, respectively). Thus, the intron 1 DHS does not seem to boost expression in the MDCK-IOWA cell line. Previous reports suggest that only gut cells upregulate expression in response to the intron 1 DHS (Smith, AN et al, 1996; Rowntree et al, 2001). This data confirms that kidney cells do not show a similar upregulation of expression, despite a high level of endogenous CFTR expression.

5.2.3 Transfection of the T84 cell line

The T84 cell line was derived from the gut, and expresses CFTR at high levels. A panel of vectors was transfected into the T84 cell line (Fig. 5.3). Once again, transfection of the pEGFP-N vector produced a high level of expression (%[f] = 11.28, X_f = 487.75), while p1kbcfproEGFP showed a lower, but still substantial level of expression (%[f] = 5.33, X_f = 221.58).

Expression from the $_{PAC}$65kbcfproEGFP vector was greatly reduced (%[f] = 0.11, X_f = 167.78) in comparison to the p1kbcfproEGFP vector. However, these results were again mirrored when comparing pEGFP-N-transfected cells to $_{PAC}$RC2cmvEGFP – transfected cells (%[f] = 0.73, X_f = 444.89). Thus in the T84 cell line, there is no evidence of control elements in the additional 5' region from –1 kb to –65kb. Again,

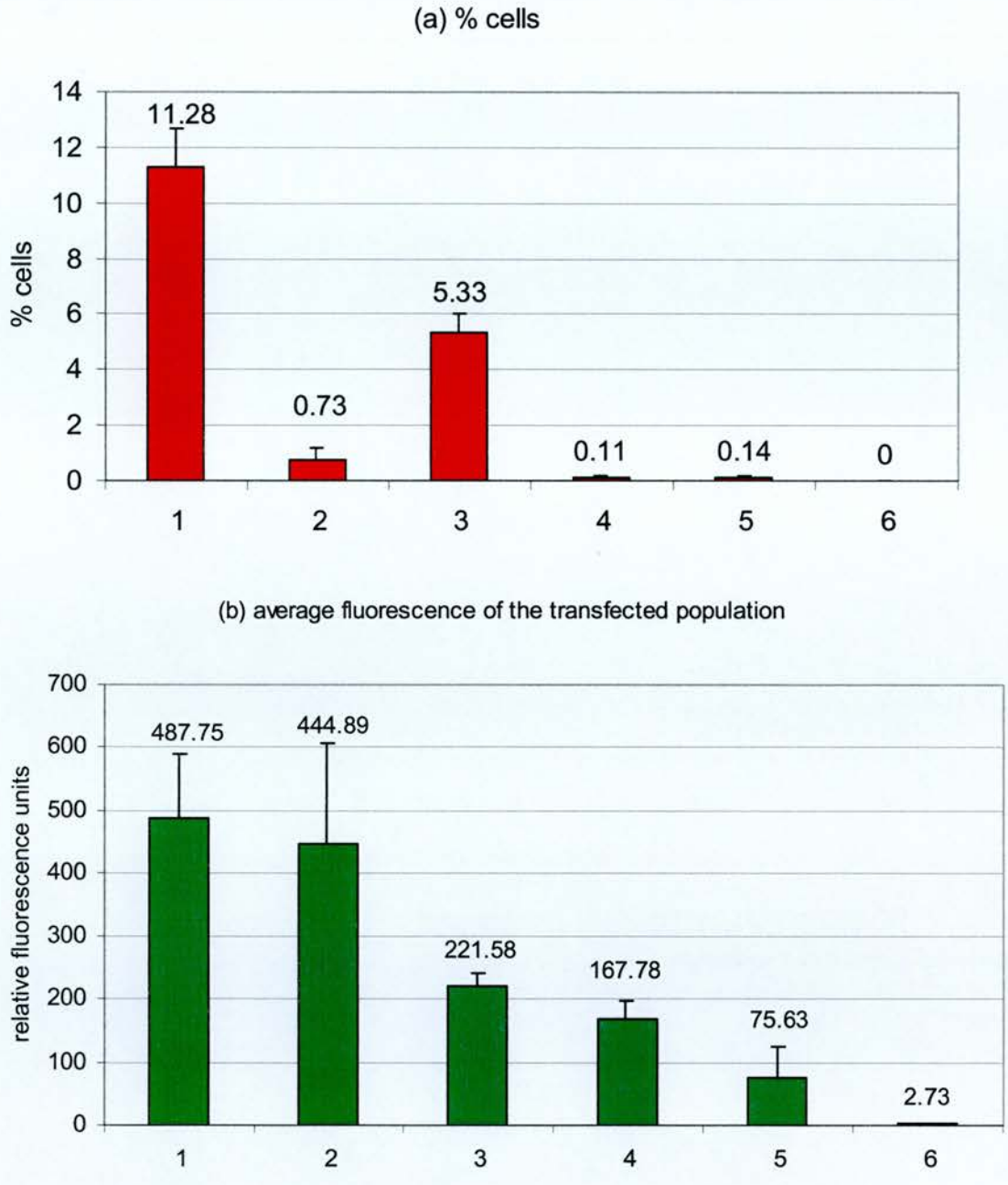


Figure 5.3. Transfection of EGFP reporter constructs into the T84 cell line. The T84 cell line was transfected with the plasmids **(1)** pEGFP-N, **(2)** $_{PAC}$RC2cmvEGFP, **(3)** p1kbcfproEGFP, **(4)** $_{PAC}$65kbcfproEGFP, **(5)** pCFTRiresEGFP, using the LID reagent, or **(6)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.11.

the promoter present in the vector had more of an effect than the vector size upon the average fluorescence of the transfected population.

No expression was detected in $_{PAC}$RC1iresEGFP or $_{PAC}$RC2iresEGFP-transfected cells (data not shown), this was probably due to the poor transfection efficiency in the T84 line. Thus, it was not possible to analyse the effects of intron 1 upon expression in this line. Transfection of the pCFTRiresEGFP vector produced few cells expressing EGFP (%[f] = 0.14) with a low average fluorescence intensity (X_f = 75.63).

5.2.4 Transfection of the Caco-2 cell line

The Caco-2 cell line was derived from a colon carcinoma. This line is frequently used in studies of CFTR expression (e.g. Smith, AN et al, 1996). The Caco-2 cell line was transfected with a panel of vectors; the results are shown in Fig. 5.4.

p1kbcfproEGFP-transfected cells produced a low level of expression (%[f] = 1.94, X_f = 234.45). This cell line is not easily transfected, and even pEGFP-N-transfected cells produced only a moderate level of expression (%[f] = 6.05, X_f = 944.74).

The differences in expression between p1kbcfproEGFP-transfected cells and $_{PAC}$65kbcfproEGFP-transfected cells (%[f] = 0.07, X_f = 69.7) roughly mirrored the difference between pEGFP-N-transfected cells and $_{PAC}$RC2cmvEGFP-transfected cells (%[f] = 0.39, X_f = 479.34). As in the MDCK-$_{IOWA}$ and T84 cell lines, the promoter present in the vector had more of an effect than the vector size upon the average fluorescence of the transfected population.

No expression was detected in $_{PAC}$RC1iresEGFP-transfected or $_{PAC}$RC2iresEGFP-transfected cells in the Caco-2 cell line (results not shown), so it was not possible to analyse the effects of the intron 1 DHS upon expression.

5.2.5 Transfection of the HBE cell line

A preliminary study was done on the HBE cell line, using the SID transfection method (see Chapter 4). Fig. 5.5 shows the results of this study.

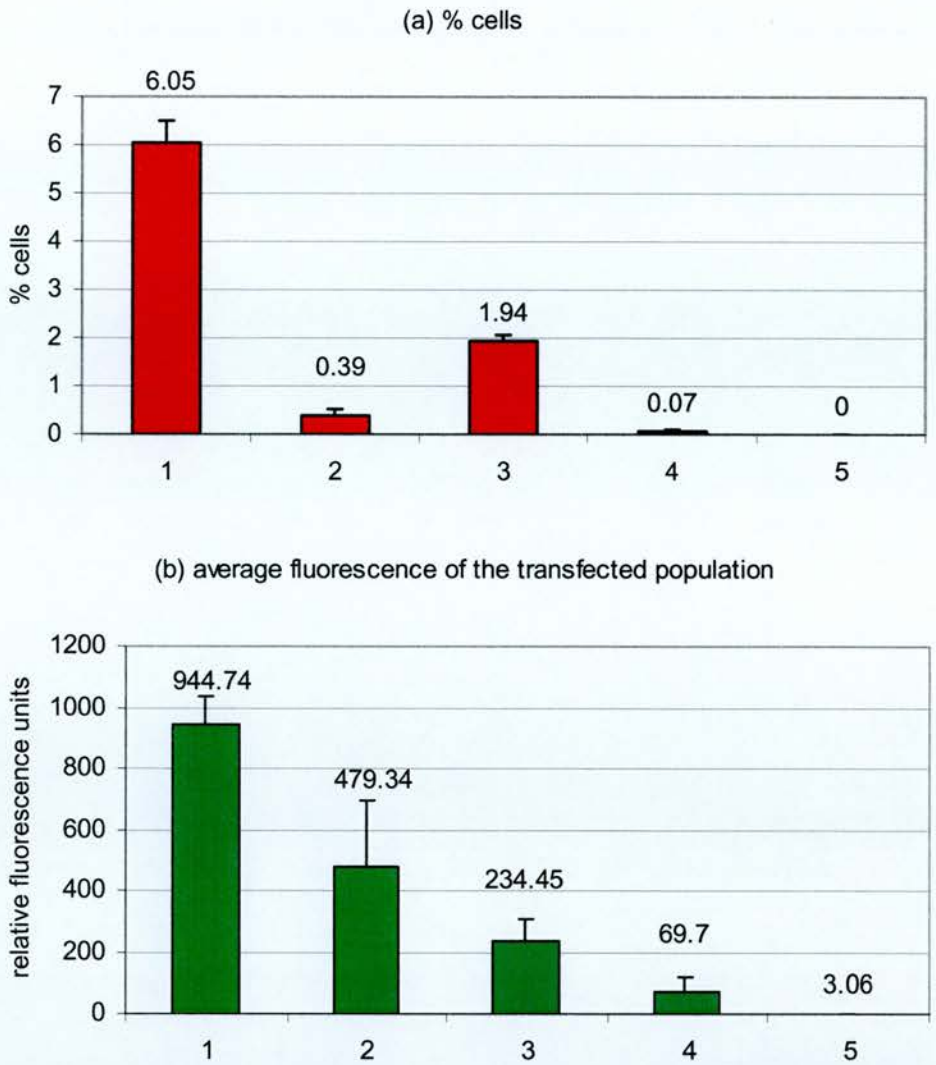


Figure 5.4. Transfection of EGFP reporter constructs into the Caco-2 cell line. The Caco-2 cell line was transfected with the plasmids **(1)** pEGFP-N, **(2)** $_{PAC}$ RC2cmvEGFP, **(3)** p1kbcfproEGFP, **(4)** $_{PAC}$65kbcfproEGFP, using the LID reagent, or **(5)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.12.

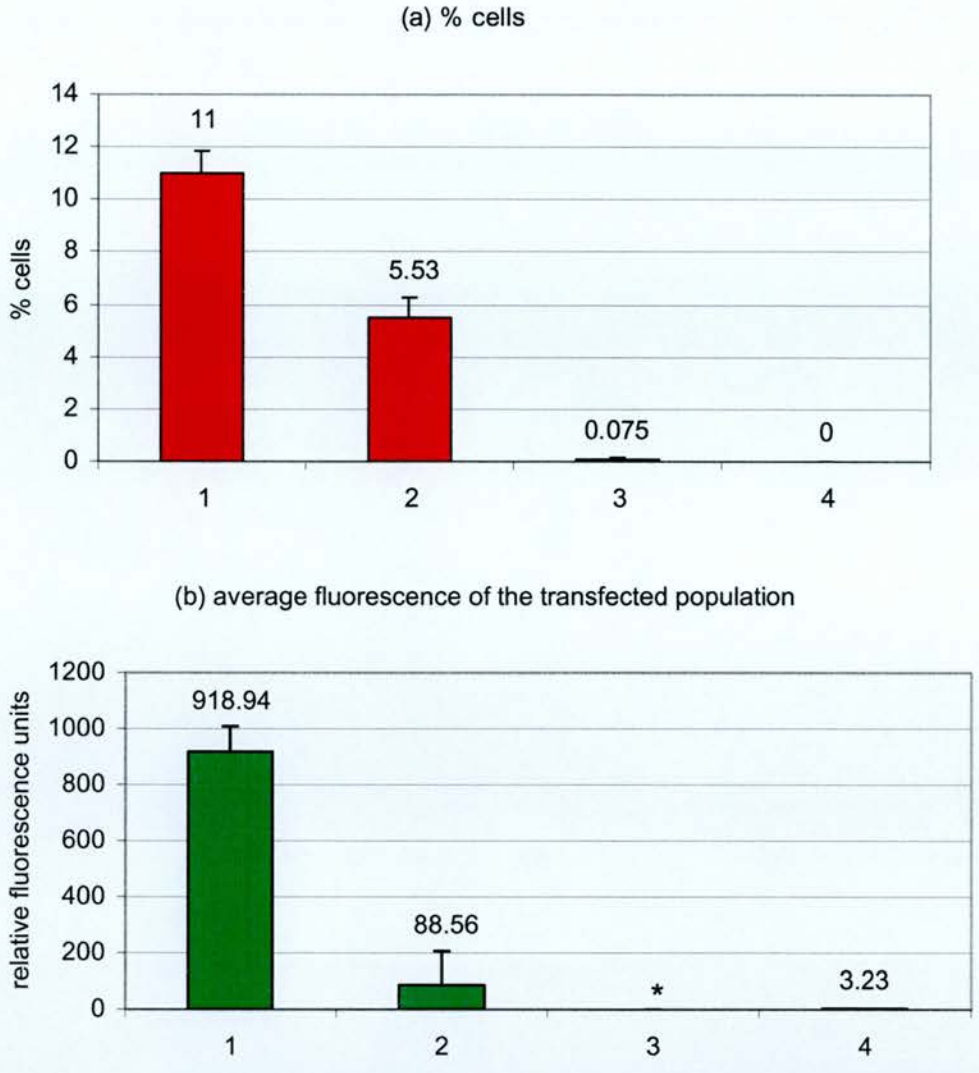


Figure 5.5. Transfection of EGFP reporter constructs into the HBE cell line. The HBE cell line was transfected with the plasmids **(1)** pEGFP-N, **(2)** p1kbcfproEGFP, **(3)** $_{PAC}$65kbcfproEGFP, using the SID reagent, or **(4)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. * the average fluorescence intensity could not be accurately calculated for the $_{PAC}$65kbcfproEGFP vector. The dotplots for these histograms are shown in Appendix C.13.

p1kbcfproEGFP-transfection produced a moderate level of expression in the HBE cell line (%[f] = 5.53%, 88.56), in contrast to a higher level of expression with pEGFP-N (%[f] = 11, X_f = 918.94). The $_{PAC}$65kbcfproEGFP vector showed a very small number of transfected cells (%[f] = 0.075). The average fluorescence intensity of this population could not be determined, as the number of transfected cells was too small to accurately sample the FL1-H intensity distribution of the transfected population. Unfortunately, time constraints and a reduction in the efficiency of the SID method prevented further experiments being conducted in this cell line during this study.

5.2.6 Evaluation of the effects of PAC backbone and vector copy number upon expression

In every cell line analysed, expression from $_{PAC}$RC2cmvEGFP was reduced in comparison to pEGFP-N, despite the fact that these two vectors contain an identical expression cassette (EGFP driven by a P_{CMV} promoter). This seems most likely to be a direct result of the difference in vector size. However, one should remain aware that the reduction in expression could be caused by other features of the PAC vector (for example, components of the PAC backbone); further experiments would be required to rule this out.

At first, we hypothesised that the reduction might be a simple effect of vector copy number. Because pEGFP-N is much smaller (with a lower molecular weight) than $_{PAC}$RC2cmvEGFP, there will be many more copies of this small plasmid in 1 µg DNA. Thus, the reduction in expression may be solely due to the delivery of a smaller number of gene copies in the PAC transfection.

An experiment was devised to test this hypothesis. 42.5 ng pEGFP-N was mixed with 957.5 ng of an anonymous plasmid - pUC18 (which does not contain a mammalian expression cassette) - to produce an equivalent number of pEGFP-N vector copies as in 1 µg $_{PAC}$RC2cmvEGFP, while maintaining the correct DNA weight (1 µg) for optimum LID complex transfection (the vectors were mixed prior to LID complex formation). This mixture was transfected into the COS-7 cell line

alongside controls of 1 μg pure pEGFP-N, 1 μg $_{PAC}$RC2cmvEGFP, and 1 μg pUC18. The FACS results for these transfections are shown in figure 5.6.

When the 42.5 ng pEGFP-N mixed with 957.5 ng pUC18 was transfected, a substantial level of expression was seen (%[f] = 32.61% and X_f = 4906.71). This was not reduced to the lower levels seen with $_{PAC}$RC2cmvEGFP transfection (%[f] = 3.65%, X_f = 441.49); surprisingly, the levels were *increased* over those seen with 1 μg pure pEGFP-N (%[f] = 22.95, X_f = 4586.43). This suggests that the reduction in expression levels from PAC vectors is not a result of the lower gene copy number present in those transfections. The profile of cells transfected with 1 μg pUC18 was similar to the untransfected control.

One might imagine several ways in which large PAC vectors could affect transfection. For example, PAC DNA might have a gross affect disturbing the transfection process (e.g. lipoplex complexes might not form as effectively in the presence of large molecules). A second experiment was done to test whether PAC vectors were able to alter transfection efficiency in trans, e.g. of a co-transfected plasmid molecule. 42.5 ng pEGFP-N was mixed with 957.5 ng $_{PAC}$RC2b (a PAC vector identical to $_{PAC}$RC2cmvEGFP, except that there is a βgalactosidase gene in place of the EGFP gene – see Appendix D). Hence, this mixture contains predominantly large PAC DNA, but the EGFP reporter gene itself is contained in a small plasmid (the vectors were mixed prior to LID complex formation). In addition, the $_{PAC}$RC2cmvEGFP PAC was mixed with $_{PAC}$RC2b in the same way, as a control. These were transfected into the COS-7 cell line alongside controls of 1 μg pure pEGFP-N, 1 μg $_{PAC}$RC2cmvEGFP, and 1 μg $_{PAC}$RC2b. The results are shown in Fig. 5.7.

When the 42.5 ng pEGFP-N mixed with 957.5 ng $_{PAC}$RC2b was transfected into the COS-7 line, there was a substantial level of expression (%[f] = 39.49, X_f = 4540.98). This was not reduced towards the levels seen with $_{PAC}$RC2cmvEGFP transfection (%[f] = 4.85, X_f = 752.19); in fact, the levels were at least as high as those in the pEGFP-N-only transfection (%[f] = 37.43, X_f = 4542.36). Thus, PAC DNA did not appear to affect the transfection efficiency of a small plasmid vector in

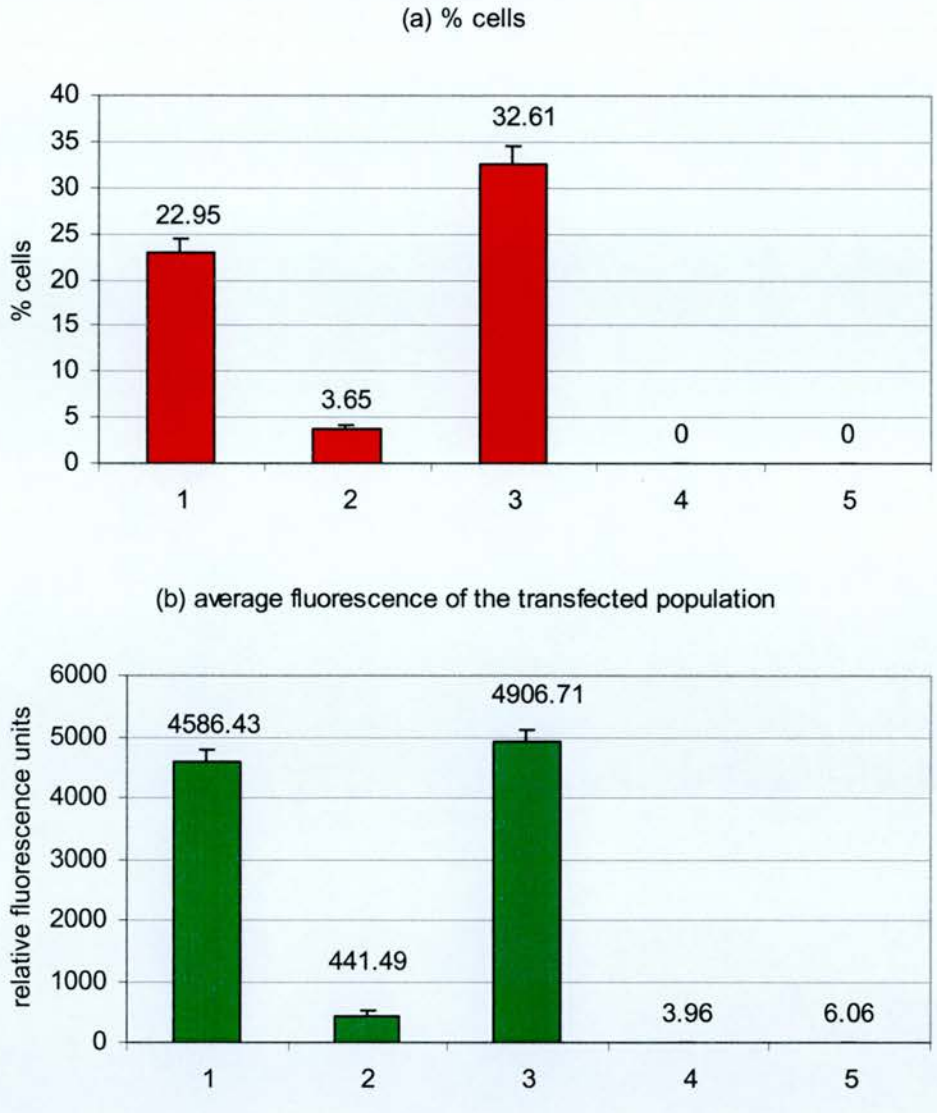


Figure 5.6.The effects of equalising pEGFP-N and pRC2cmvEGFP vector copy number. The COS-7 cell line was transfected with **(1)** 1μg pEGFP-N, **(2)** 1 μg $_{PAC}$RC2cmvEGFP, **(3)** 42.5 ng pEGFP-N + 957.5 ng pUC18, **(4)**1 μg pUC18 using the LID reagent, or **(5)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.14.

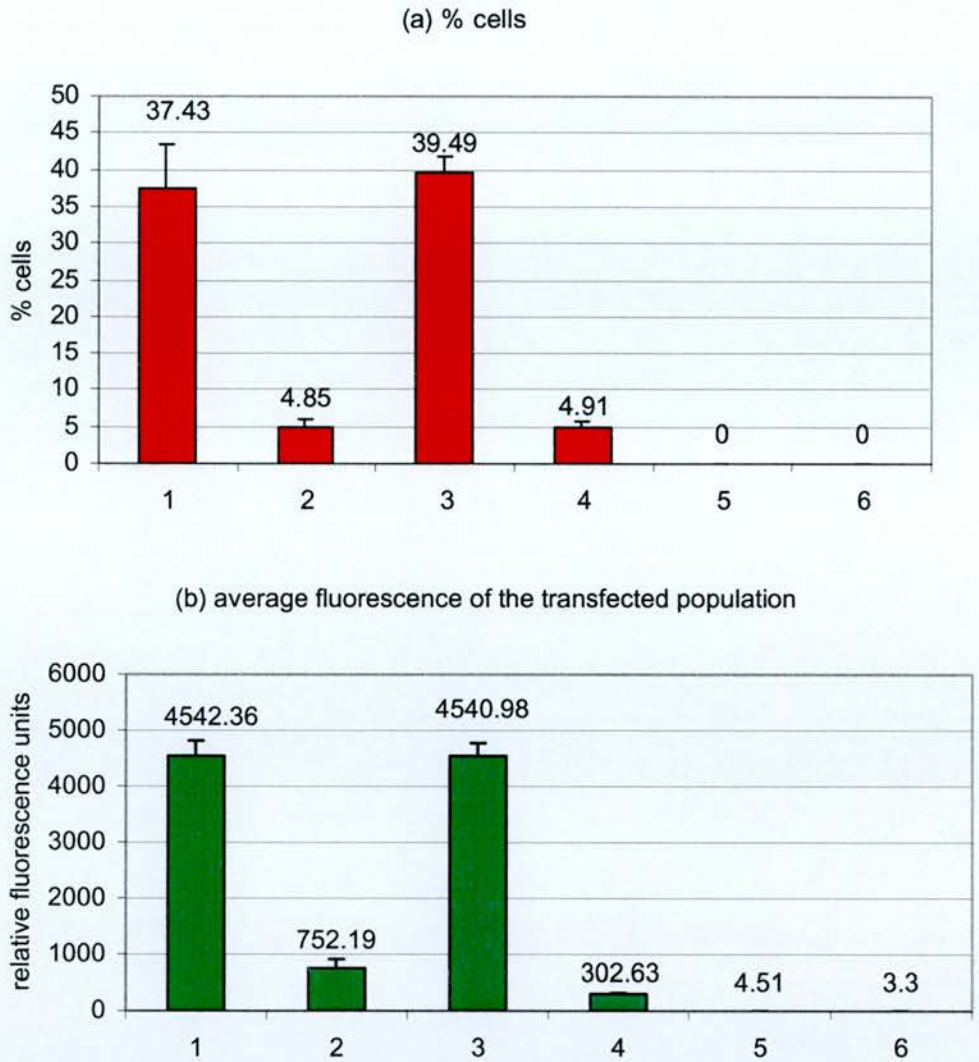


Figure 5.7.The effects of diluting pEGFP-N with $_{PAC}$RC2b. The COS-7 cell line was transfected with **(1)** 1μg pEGFP-N, **(2)** 1 μg $_{PAC}$RC2cmvEGFP, **(3)** 42.5 ng pEGFP-N + 957.5 ng $_{PAC}$RC2b, **(4)** 42.5 ng $_{PAC}$RC2cmvEGFP + 957.5 ng $_{PAC}$RC2b **(5)** 1 μg $_{PAC}$RC2b using the LID reagent, or **(6)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.15.

trans. The reduced expression of cassettes in cis seems to be a direct effect of the overall size of the DNA molecule. Transfection of the $_{PAC}$RC2b vector alone produced a similar profile to the untransfected sample.

The fact that transfection efficiency was not reduced following a 1/24 dilution with anonymous pUC18 DNA was in itself surprising, and some experiments were done to investigate this further. pEGFP-N was serially diluted with pUC18 and transfected into the COS-7 cell line, using the LID reagent (Fig. 5.8). A curve of transfection efficiency was seen following dilution. Transfection of pure pEGFP-N produced only a moderate level of expression in this experiment (%[f] = 8.86, X_f = 3774.09) on this day. Surprisingly, dilution of the plasmid 1/10 or 1/100 with pUC18 seemed to increase the percent of cells expressing EGFP (%[f] = 15.6 and 10.5 respectively), without significantly affecting the average fluorescence intensity of the transfected population (X_f = 3858.18 and 3024.92, respectively). The profile of cells transfected with 1 µg pUC18 was similar to the untransfected control.

Next, we tested the effects of diluting a PAC reporter construct. $_{PAC}$RC2cmvEGFP was serially diluted with pUC18 and transfected into the COS-7 cell line using the LID reagent (Fig. 5.9). Again, a curve of transfection efficiency was seen. Transfection of pEGFP-N produced a substantial level of expression (%[f] = 20.04, X_f = 4432.6), while pure $_{PAC}$RC2cmvEGFP produced a low level of expression (%[f] = 2.86, X_f = 642.71). The percentage of cells expressing EGFP was increased upon diluting the $_{PAC}$RC2cmvEGFP 1/10 or 1/100 with pUC18 (%[f] = 6.25 and 3.84, respectively). The average fluorescence intensity was not significantly affected in the 1/10 dilution (X_f = 605.26), although there was a slight drop in the average fluorescence intensity in the 1/100 dilution (X_f = 333.24). Again, transfection of pUC18 alone produced a similar profile to the untransfected control. Thus, the mixing effect is not specific to the pEGFP-N vector, and can increase transfection efficiency for a PAC vector, too. Looking back at Fig. 5.7, we also find evidence that we may be able to dilute a reporter PAC with a non-coding PAC, without compromising its transfection efficiency. Mixing 42.5 ng $_{PAC}$RC2cmvEGFP with 957.5 ng $_{PAC}$RC2b did not reduce the percentage of cells expressing EGFP (%[f] = 4.91), in comparison with 1 µg pure $_{PAC}$RC2cmvEGFP (%[f] = 4.85), although the

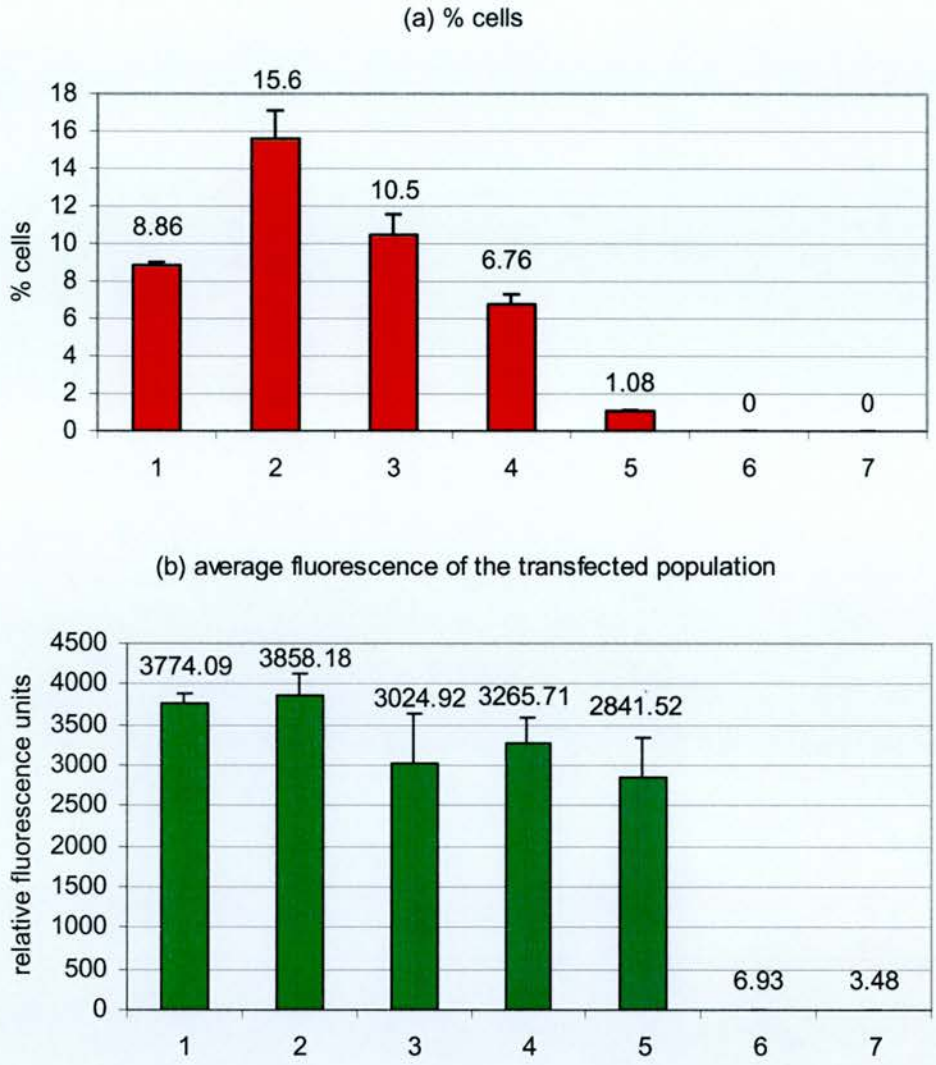


Figure 5.8.The effects of serial dilution of pEGFP-N with pUC18. The COS-7 cell line was transfected with **(1)** 1μg pEGFP-N, **(2)** 100 ng pEGFP-N + 900 ng pUC18, **(3)** 10 ng pEGFP-N + 990 ng pUC18, **(4)** 1 ng pEGFP-N + 999 ng pUC18 **(5)** 100 pg pEGFP-N + 999.9 ng pUC18, **(6)** 1 μg pUC18 using the LID reagent, or **(7)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.16.

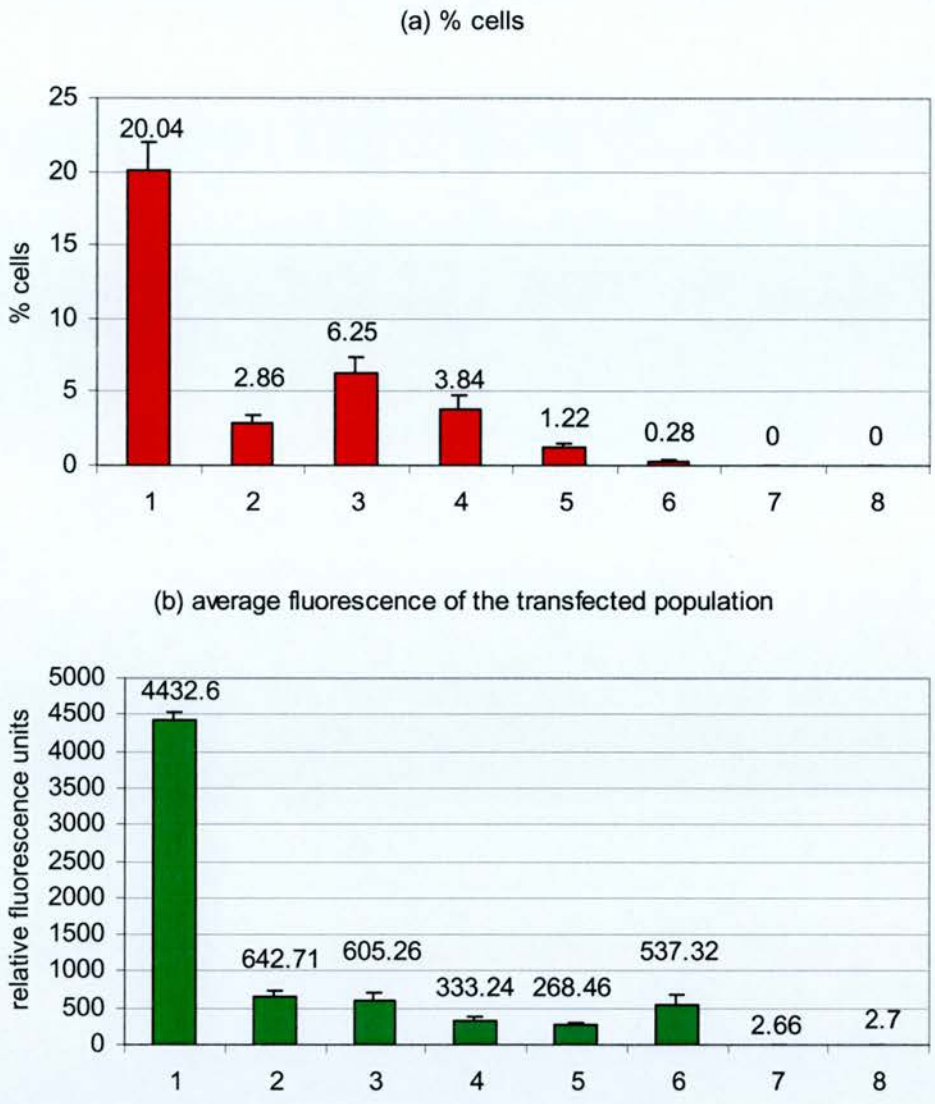


Figure 5.9.The effects of serial dilution of $_{PAC}$RC2cmvEGFP with pUC18. The COS-7 cell line was transfected with **(1)** 1μg pEGFP-N, **(2)** 1 μg $_{PAC}$RC2cmvEGFP **(3)** 100 ng $_{PAC}$RC2cmvEGFP + 900 ng pUC18, **(4)** 10 ng $_{PAC}$RC2cmvEGFP + 990 ng pUC18, **(5)** 1 ng $_{PAC}$RC2cmvEGFP + 999 ng pUC18 **(6)** 100 pg $_{PAC}$RC2cmvEGFP + 999.9 ng pUC18, **(7)** 1 μg pUC18 using the LID reagent, or **(8)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.17.

average fluorescence of the transfected population was slightly reduced (X_f = 302.63 vs. 752.19).

Subsequently, we tested whether it was necessary for the anonymous DNA to be present in the same complex as the pEGFP-N reporter to produce an increase in transfection efficiency. 100 ng pEGFP-N was either mixed with 900 ng pUC18 prior to LID complex formation, or appropriate amounts of lipofectin and P6 were added to each plasmid to form separate complexes, which were then applied simultaneously to a sample of COS-7 cells for transfection. In order to rule out the possibility that it is merely a reduction in the amount of reporter DNA, rather than the inclusion of an anonymous DNA plasmid, which enhances transfection, two additional controls were included. In the first control, the amount of DNA within the LID complex was reduced to 100 ng, while the quantities of the other LID components were kept constant (0.75 µl lipofectin and 4 µg P6). In the second control, the entire LID complex was scaled down (thus 100 ng DNA was complexed with 0.075 µl lipofectin and 0.4 µg P6). These were transfected into the COS-7 cell line, alongside controls of 1 µg pEGFP-N and 1 µg pUC18 (Fig. 5.10).

Once again, mixing the pEGFP-N plasmid with pUC18 plasmid prior to complexing, increased expression (%[f] = 32.91, $X_{f=}$ 4626.76), in comparison to transfection of pure pEGFP-N plasmid (%(f) = 29.46, X_f 4= 4543.98). When the plasmids were complexed separately and then applied to the cells simultaneously for transfection, the expression was reduced (%[f] = 13.85, X_f = 3908.96). This shows that it is necessary to mix anonymous DNA with the reporter plasmid before complex formation (so that the two plasmids are present in the same complex) in order to achieve an increase in transfection efficiency.

In addition, expression was reduced in transfections where the DNA quantity was reduced within the complex (%[f] = 12.04, X_f = 3889.25), or where the entire complex quantity was reduced (%[f] = 0.33, X_f = 4168.25). This shows that the increase was not an indiscriminate effect of lowering the transgene copy number.

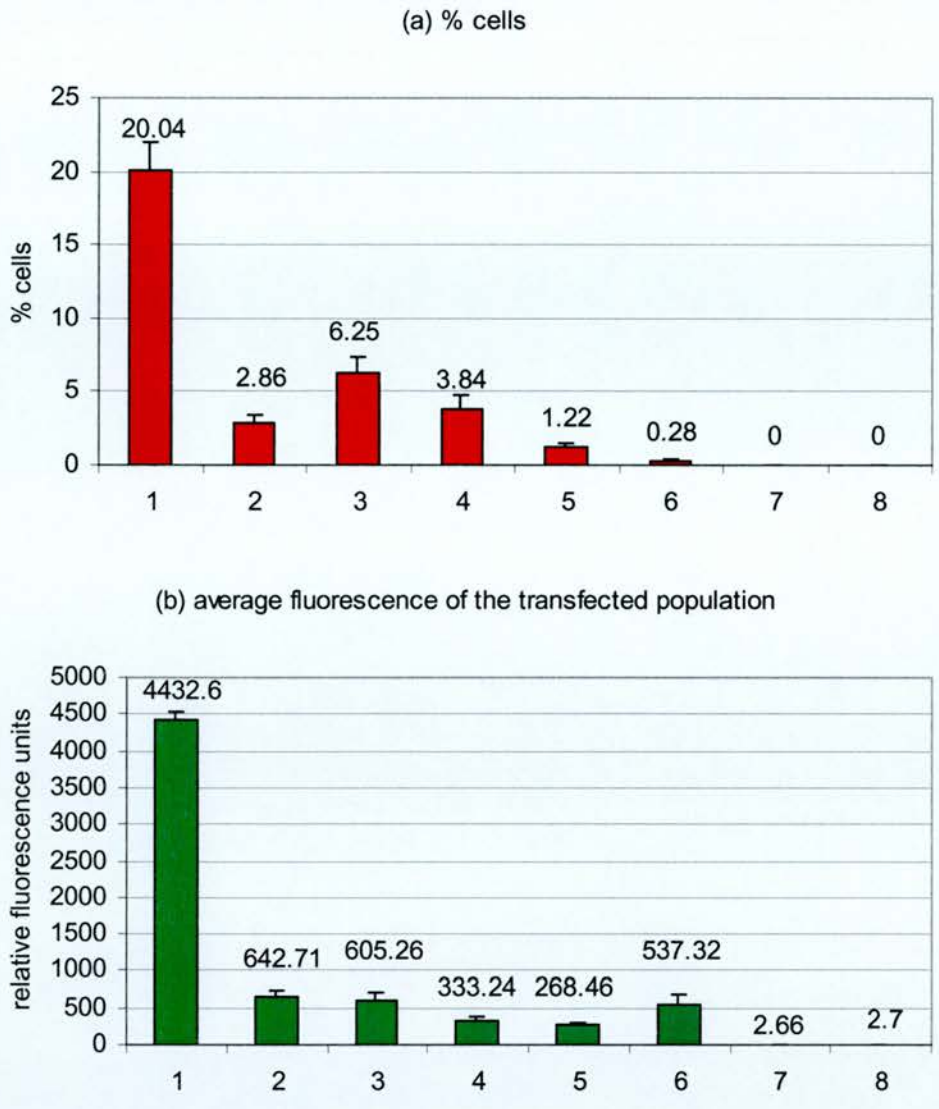


Figure 5.9.The effects of serial dilution of $_{PAC}$RC2cmvEGFP with pUC18. The COS-7 cell line was transfected with **(1)** 1µg pEGFP-N, **(2)** 1 µg $_{PAC}$RC2cmvEGFP **(3)** 100 ng $_{PAC}$RC2cmvEGFP + 900 ng pUC18, **(4)** 10 ng $_{PAC}$RC2cmvEGFP + 990 ng pUC18, **(5)** 1 ng $_{PAC}$RC2cmvEGFP + 999 ng pUC18 **(6)** 100 pg $_{PAC}$RC2cmvEGFP + 999.9 ng pUC18, **(7)** 1 µg pUC18 using the LID reagent, or **(8)** left untransfected. 48 hours after transfection, cells were analysed by FACS (see Methods). This figure shows: **(a)** a histogram of the percentage of cells with detectable EGFP expression and **(b)** a histogram of the average fluorescence of the transfected population. The dotplots for these histograms are shown in Appendix C.17.

5.3 Discussion

5.3.1 CFTR regulatory elements

These studies show that the proximal CFTR 5' region is capable of driving EGFP expression in a range of cell lines. In every cell line tested, p1kbcfproEGFP transfections produced a lower average fluorescence intensity amongst the transfected population, in comparison to pEGFP-N transfections. The percentage of cells expressing EGFP was also lower in every cell line except COS-7 (where they were approximately equal). Thus, the results of this study are consistent with the hypothesis that the CFTR proximal 5' region has weak promoter activity, reported by others (Yoshimura et al, 1991b), in contrast to the strong P_{CMV} promoter. The data presented here are consistent with the proximal promoter region retaining this characteristic in these transient transfection systems.

The $_{PAC}$65kbcfproEGFP transfections showed a reduction in both the percentage of cells expressing EGFP and the average fluorescence of the transfected population, in comparison to p1kbcfproEGFP transfections, in every cell line tested. This result was mirrored when comparing $_{PAC}$RC2cmvEGFP (large PAC) and pEGFP-N (small plasmid) transfections, despite the fact that these vectors contain the same promoter (P_{CMV}). Thus, it seems that a reduction in expression is intrinsic to large PAC vectors. In light of this effect, any positive or negative modulation of expression *per se* by the additional upstream region in the $_{PAC}$65kbcfproEGFP would be difficult to discern.

An attempt was made to analyse whether the difference in the average fluorescence intensity of the p1kbcfproEGFP (μ1) and $_{PAC}$65kbcfproEGFP (μ2) vectors was significantly different to the difference in average fluorescence intensity of the pEGFP-N (μ3) and $_{PAC}$RC2cmvEGFP (μ4) vectors, in the cell lines COS-7, MDCK-$_{IOWA}$, T84 and CaCO2. The data for the transfected populations was extracted and converted from logarithmic to linear format using the Excel software package, for this purpose. Then, a statistical analysis was performed by Naomi Wray, MMC, University of Edinburgh.

A generalised linear model was used (treating the triplicate transfections as a random effect). We want to test the hypothesis $\mu1/\mu2 = \mu3/\mu4$. However, this hypothesis is non-trivial. The hypothesis $\mu1\mu4 = \mu2\mu3$ is equivalent, and easier to test. The results of this analysis are shown in Appendix F2.

In summary, although $\mu1\mu3$ and $\mu2\mu4$ were very similar (a difference of 0.1, or less), this difference was found to be statistically significant in every cell line. The sign of the test statistic (negative or positive) was not consistent between cell lines (relating to whether $\mu1\mu4$ or $\mu2\mu3$ is greater). The fact that the differences were very slight and the inconsistency in the statistic sign suggest that the difference was not due to an enhancer effect of the additional region present in the $_{PAC}$65kbcfproEGFP vector.

It is difficult to interpret the results of the statistical analysis in terms of the experiment. Because the sample size is so large (10,000 cells or greater), we can calculate $\mu1\mu4$ and $\mu2\mu3$ very accurately. The significant difference that we see is likely to be due to experimental artefacts, or small differences in vector structure (e.g. p1kbcfproEGFP and pEGFP-N are of similar, but not identical size; the same is true of the $_{PAC}$RC2cmvEGFP and $_{PAC}$65kbcfproEGFP vectors), rather than regulatory elements.

Ann Harris' group has performed some elegant experiments testing the effects of the –20.9 DHS upon expression. Briefly, two YAC vectors were made containing the CFTR locus. The first YAC included ~50 kb of CFTR 5'region (Vassaux et al, 1997). In the second, truncated YAC the –20.9 DHS and all upstream DNA was deleted (Nuthall et al, 1999b). Permanent Caco-2 cell lines were made with these two YACs; expression of YAC-derived CFTR mRNA appeared to be copy-number dependent, position independent in these lines. In the two permanent lines made from the intact YAC, the ratio of YAC-derived CFTR mRNA to endogenous CFTR mRNA was proportional to the ratio of YAC copies to endogenous CFTR copies. However, in the five permanent lines made from the truncated YAC, less CFTR mRNA was produced for each YAC copy than for each endogenous CFTR copy. Semi-quantitative RT-PCR suggested that expression was reduced by ~60% in each

YAC copy, in comparison to each endogenous copy. Thus, there is convincing evidence that the –20.9 DHS can act as a positive regulatory element.

This experiment has not been repeated in other cell lines, so it is not known whether this upregulation is specific to the Caco-2 cell line, or a more general phenomenon. A previous report (Smith et al, 1995) analysed the presence of 5' DHS elements in a panel of cell lines. The –20.9 DHS was found to be present in every cell line analysed, including those that did not express CFTR, and the level of DNase hypersensitivity was not correlated with the level of CFTR expression in these lines. Hence, the –20.9 DHS appears to be a ubiquitous, rather than tissue-specific, regulatory element.

Two PMR elements (regions of purine-pyrimidine strand asymmetry) co-localise with the –20.9 DHS. These elements may play a role in the upregulation of expression (see Section 1.1.7). While the PMR elements at –20.9 were shown not to adopt non-B-DNA conformation; it is still possible that they will affect expression by altering the chromatin structure in this region. There is additional evidence that chromatin conformation may be involved in transcriptional regulation at the CFTR locus (Li et al, 1999, Hollingsworth et al, 1994). It is unclear whether histones become assembled upon transiently transfected DNA. CpG methylation may also be necessary for proper transcriptional regulation, and this may not be accomplished in transiently transfected DNA. Hence, genomic integration might be required for some of the CFTR regulatory elements (for example, the –20.9 DHS) to elicit an effect upon expression.

One criticism that could be made of the Nuthall et al, 1999 paper is that a semi-quantitative technique (phosphoimager detection of ^{32}P signal in an RT-PCR Southern Blot hybridisation) was used to measure the levels of CFTR expression. As they have not included quantitative controls, it is difficult to know whether the 60% reduction in the phosphoimager signal per YAC copy (in relation to each endogenous CFTR copy) is truly representative of a 60% reduction in mRNA levels; in reality, the effect may be smaller or greater. While there is likely to be a relationship between the phosphoimager signal and the cDNA copy number, the ratio is unlikely to be 1:1. A quantitative technique, such as the TaqMan RT-PCR assay (Heid et al,

1996), or primer extension analysis of mRNA (Walmsley et al, 1998) would provide a more accurate measure of the mRNA levels (Carey and Smale, *Transcriptional Regulation in Eukaryotes*, Chapter 5).

Certainly, modulation of expression by the –20.9 DHS appears to be subtle (for example, in relationship to the large differences seen between the strong P_{CMV} promoter and weak CFTR promoter). Because the effect of vector size upon expression has proved to be so extreme, it would be difficult to analyse whether the $_{PAC}$65kbcfproEGFP exhibits this subtle upregulation of expression. The data presented here cannot distinguish between the hypothesis that the additional region has no effect, or a small positive effect, upon expression of the EGFP transgene.

When the $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP vectors were transfected into the COS-7 and MDCK-IOWA cell lines, they produced similar levels of expression. Thus, it appears that the intron 1 element does not modulate expression in either of these lines. It is unfortunate that poor transfection efficiency in the T84 and Caco-2 cell lines prevented an analysis of the gut-cell specific enhancement of expression by the intron 1 element. These cell lines were generally difficult to transfect; when coupled with the effects of a weak CFTR promoter, large PAC context, and inefficient translation from an IRES, it is perhaps not surprising that no expression was detected from the $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP constructs.

The efficiency of transient transfection in the T84 and Caco-2 cell lines could probably be further improved by novel transfection methods. Alternatively, it might be possible to generate stably transfected cell lines, or even transgenic mice, to address this question (although this would no longer be an accurate model for the effects of these vectors in a transient transfection/gene therapy context).

Previous reports (Smith, AN et al, 1996; Mogayzel and Ashlock, 2000, Rowntree et al, 2001) have shown that the intron 1 DHS element upregulates expression in a tissue specific fashion, in gut epithelial cells that express CFTR. This element can successfully increase expression from the CFTR proximal promoter in reporter constructs, even in a transient transfection assay. In other words, while the –20.9 DHS may require genomic integration to elicit an effect upon expression, it seems

unlikely that this will be required for the intron 1 DHS to affect expression. Hence, the intron 1 element might be capable of modulating expression in gene therapy vectors.

The results presented here confirm that the intron 1 element does not upregulate expression in a cell line that does not express CFTR (COS-7) or in a cell line that expresses CFTR at high levels, but does not derive from the gut (MDCK-IOWA). These results are consistent with the hypothesis that this element is specific to gut-type cells expressing CFTR. If this proves to be true, the intron 1 DHS may ultimately be irrelevant to CF gene therapy vectors, as this therapy is likely to focus on the respiratory tissues. The high levels of CFTR expression in some tissues outside the gut (e.g. the kidney tubules, pancreatic ducts and tracheal submucosal glands), suggests that there must be other tissue-specific regulatory elements at the CFTR locus, in addition to the intron 1 DHS.

CFTR is expressed at high levels in only a small number of cells in both the respiratory tract - a subpopulation of cells in the submucosal gland (Engelhardt et al, 1992), and the gut – ductal cells of the intestinal crypts (Trezise and Buchwald, 1991; Crawford et al, 1991). It may be necessary to produce cell lines that approximate these cell types, to see tissue-specific upregulation of expression. It is also possible that *in vivo* organ structures are required to stimulate CFTR expression in these populations of cells. Despite a multitude of studies investigating the endogenous CFTR location, the identity of CFTR-expressing cells has not been precisely elucidated. Other groups have encountered similar difficulties in dissecting CFTR regulatory elements, according to their effect on a LacZ transgene (Imler et al, 1996; Chou et al, 1991).

5.3.2 The effects of vector size and plasmid mixing upon expression

In every cell line analysed, the vector size (e.g. plasmid or PAC), had a large effect on the percentage of cells expressing EGFP, while the promoter species (e.g. P_{CMV} or CFTR 5' region) had a smaller effect. However, in three out of four cell lines analysed, the promoter had the major effect upon the average fluorescence of the transfected population, while the vector type had a lesser effect. The exception to this

rule was COS-7, where the vector type had a larger effect upon the average fluorescence of the transfected population. This suggests that the COS-7 cell line is atypical, and may have been artificially selected to promote transfection of small plasmids (but does not accomplish this with large plasmids). Although experiments in the COS-7 line paint a bleak picture of PAC construct transfection, this does not seem to be a typical result, and may not be reflected *in vivo*.

Although transfection is used in a wide range of biological applications, many facets of this process are poorly understood. This would certainly not be the first time that PAC vectors had been observed to exhibit reduced expression levels in comparison to small plasmids, but little has been done to investigate the specifics, or origins, of this effect.

Preliminary experiments comparing the $_{PAC}$RC2b and pCMVβ βgalactosidase-expressing vectors confirmed that transfection efficiency was reduced for large PAC vectors, in comparison to small plasmid vectors (results not shown). One might expect that this be solely due to a reduction in the numbers of cells transfected with the large PAC vector. FACS analysis of EGFP expression has allowed us to dissect the components of transfection. This report showed that the reduction in PAC expression is a combination of a reduction in both the numbers of cells expressing the transgene and the level of expression per cell. This study showed that vector copy number is not responsible for this effect, and that the effect cannot act in trans (e.g. upon a co-transfected small plasmid).

It would be interesting to test whether this effect is a feature of large vector size *per se* or if specific sequences within the PAC reduce expression. Studies on other large vectors, such as BACs, and small plasmids incorporating PAC backbone elements could address this issue.

One hypothesis is that the larger PAC vectors are less efficiently transferred to the nucleus after cell entry. We would expect this to produce a large reduction in the average fluorescence of the transfected population (as less copies of the gene would get into the nucleus in every cell that has taken up DNA). However, we would expect this to have only a smaller, secondary effect in reducing the percentage of cells

transfected (in cases where the number of copies within the nucleus is lowered to such extent that the amount of EGFP produced is reduced below the threshold of detectability). In fact, %[f] and X_f are reduced by the same magnitude. Thus, while differential nuclear translocation is an attractive hypothesis and may be a compounding factor, it is not sufficient to explain these results alone.

It was very surprising to discover that dilution of plasmid DNA does not compromise transfection efficiency: in fact, a dilution of up to 1/100 *enhanced* transfection. Although we cannot explain this effect, there are a number of potential applications.

In applications where expensive, high quality (e.g. endotoxin-free) DNA must be used, it might be cost-effective to generate a large bulk of a generic plasmid, which can then be mixed with smaller amounts of a custom-made plasmid. Many studies (especially *in vivo* studies) are limited by sample number; it may be possible to mix multiple reporter constructs without compromising the efficiency of transfection, allowing multiple readings to be done on a single sample. It would be interesting to test whether mixing the reporter construct with other types of DNA, for example genomic or CpG methylated DNA, could increase the transfection efficiency even further. It will be necessary to test whether these effects of plasmid mixing extend to *in vivo* applications.

These plasmid-mixing experiments would certainly provide a convenient explanation for one report. Braun and colleagues (Braun et al, 1998) performed a trial investigating the possibilities of mixing up to four plasmids encoding different viral membrane bound glycoproteins for DNA immunization: they found that that mixing the plasmids had very little effect upon the magnitude or bias of the immune response to the individual components.

In summary, these studies investigated the mechanics of CFTR transcriptional regulation and the possibility of recapitulating such regulation in gene therapy vectors. The proximal 5' region was capable of recapitulating the activity of the CFTR basal promoter in these transfection studies and may be suitable for use in gene therapy vectors. Further studies will be needed to fully assess the influence of the –20.9 and intron 1 DHS elements upon expression in a gene therapy context.

Chapter 6: Primary tracheal air-interface cultures

6.1 Introduction

Although permanent cell lines form a good starting point for the analysis of gene therapy vectors, these have proved not to be very good models of gene transfer *in vivo*. One avenue for testing genomic context vectors would be animal experiments. However, the cost and time scale of such an approach is restrictive. An *ex vivo* model for analysing genomic context vectors should provide an alternative, but more realistic model for gene therapy.

Several *ex vivo* culture systems have been developed to model the respiratory system. For example, slices of whole lung tissue from mouse and humans have been successfully cultured on a membrane at air interface (McBride et al, 2000), and intact tracheae have been cultured *ex vivo* (Scott et al, 2000). In a bronchial xenograft model (Zhang et al, 1998b) human bronchial cells are seeded into denuded rat tracheae, which are then implanted into the flanks of mice; this model recapitulates many features of the native tissue, such as the cylindrical airway shape. In a simpler model, several groups have dissociated cells from the bronchus or trachea of mice (Davidson et al, 2000) humans (Gruenert et al, 1995; Zhang et al, 2001), or other species (Kondo et al, 1997), and cultured these cells on a membrane at air-interface. Although these cultures do not maintain the original tissue architecture, the cells appear to differentiate to form a polarised, pseudostratified culture, with many features of the native epithleium.

For this study, the primary culture of tracheal cells at air-interface was selected. In this method, the trachea is dissected and cells are dissociated. Fibroblast and red blood cells are selectively removed, leaving a suspension of predominantly epithelial cells. The cells are then seeded on a semi-permeable membrane and cultured at air-interface.

Initially, both sheep and mouse tracheal cultures were attempted, but the sheep was eventually chosen for this *ex vivo* study, because the human respiratory system is more similar to that of the sheep than to that of the mouse (see section 1.1.12 for a comparison of the mouse and sheep as model organisms for CF gene therapy). Another advantage of using the sheep over the mouse for these primary cultures is

their greater tissue volume. Thus, while the cells from two mouse tracheae must be combined to generate a single air-interface culture well, the cells from a single sheep trachea can generate up to 48 culture wells. This serves to reduce the number of animals required for experiments and increases sample homogeneity.

This culture method has been developed in concert with A. Doherty, working on a separate project. Results were combined and crosschecked in order to decide upon the best methods and protocols, and several aspects (such as electron microscopy) were performed in tandem for the two projects. However, the data presented here is specifically from this PhD project.

The culture architecture was characterised by scanning and transmission electron microscopy. This was done over a four-week time course, at one-week intervals, to record the development of the cultures and select the time-point that most closely mimics the *in vivo* structure. The electrical resistance across the culture was recorded to monitor the formation of tight junctions. Visual examination under the light microscope was performed routinely.

Transfection was optimised at a number of time points. Finally, some of the EGFP reporter constructs (see Chapter 3, and fold-out Appendix E) were introduced into these cultures. The FACS dotplots from which these calculations are derived are shown in Appendix C.

6.2 Results

6.2.1 Establishing Primary Cultures

Tracheae were obtained from the Moredun Research Institute: primary cultures were established as descried in the Materials and Methods Section. Briefly, tracheae were dissected and the cells were dissociated overnight. Red blood cells were removed from the suspension by preferential lysis (see Methods). This step was added to the established protocol because these were found to be a major contaminating cell type. Fibroblasts were then removed by selective adherence, leaving a suspension containing predominantly epithelial cells. Cells were seeded on the semi-permeable membrane of a Costar Transwell insert that had been pre-coated with collagen. 24

hours after seeding, the media above the insert was removed, leaving only the media below the membrane to feed the cells: this creates an air-interface, mimicking that of the airway surface. Cells were cultured for up to five weeks. This method was adapted from the protocol by L. Hyndman and H. Davidson-Smith; 'Batch 0' was set up with the kind help of L. Hyndman.

The number of cells obtained from a single trachea varied between experiments, sufficing to produce between eighteen and forty-eight cultures. The quantity of cells obtained was a function of trachea size and other factors. For example, the centrifuge used for spinning down the cells from suspension, and the length of time allowed for fibroblast adhesion affected the harvesting efficiency.

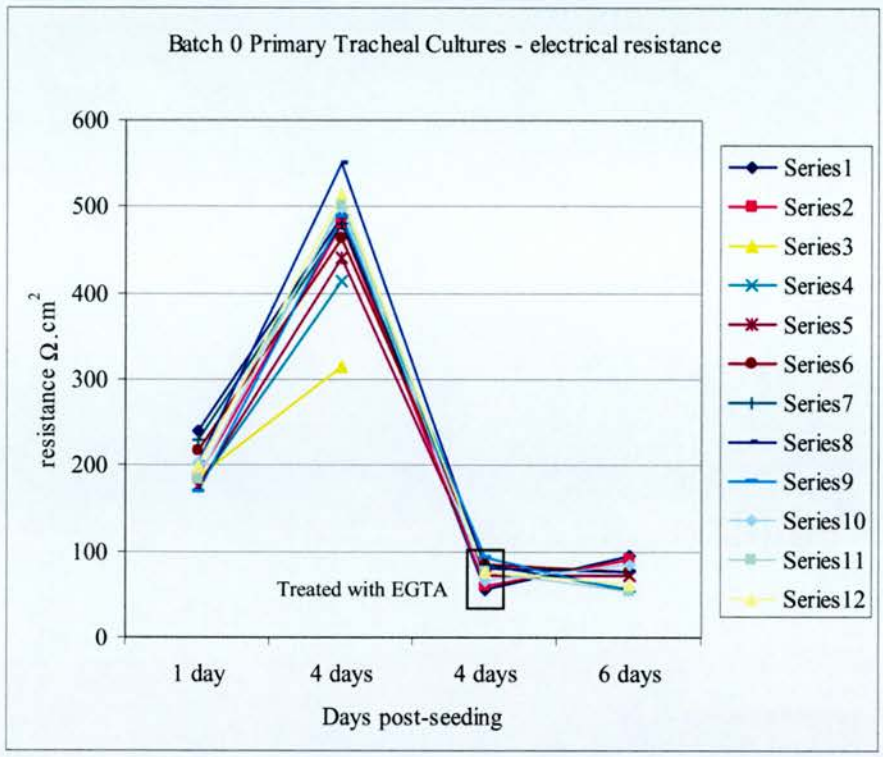
6.2.2 Electrical Resistance

The electrical resistance across the culture was measured with a voltohmeter, in order to monitor the formation of tight junctions (see Methods). Tight junctions, connecting adjacent cells, prevent an electrical current from flowing through gaps between the cells. Thus in the presence of tight junctions, the electrical current must travel transepithelially, resulting in a high resistance. As the insert has an area of 0.3 cm^2, the original readings were multiplied by 0.3 to calculate the resistance in $\Omega.cm^2$. Fig. 6.1 shows the series of resistance readings for each batch of primary cultures (batches are numbered 0- 13).

Resistance readings varied greatly, both between and within batches of cultures (see Fig. 6.1). A reading of >300 $\Omega.cm^2$ was considered indicative of tight junctions (personal communications from Donald Davidson and Hazel Davidson-Smith), and where possible, experiments were only performed on cultures that exhibited resistances of >300 $\Omega.cm^2$.

Before transfection, cultures were treated for 20 minutes with 6 mM EGTA, to disrupt tight junctions (this is required for successful transfection of these cultures – Coyne et al, 2000; Cohen et al, 2001). Transepithelial resistance was used to monitor the disruption of tight junctions; EGTA treatment resulted in a drop in resistance to below 166 $\Omega.cm^2$ (often as low as 30-60 $\Omega.cm^2$).

Figure 6.1. Primary tracheal culture resistance readings. This figure shows the resistance readings for primary tracheal cultures batches 0-13 (as labelled). Each series shows the resistance of the culture of corresponding number at different time points, and with EGTA treatment (as labelled)



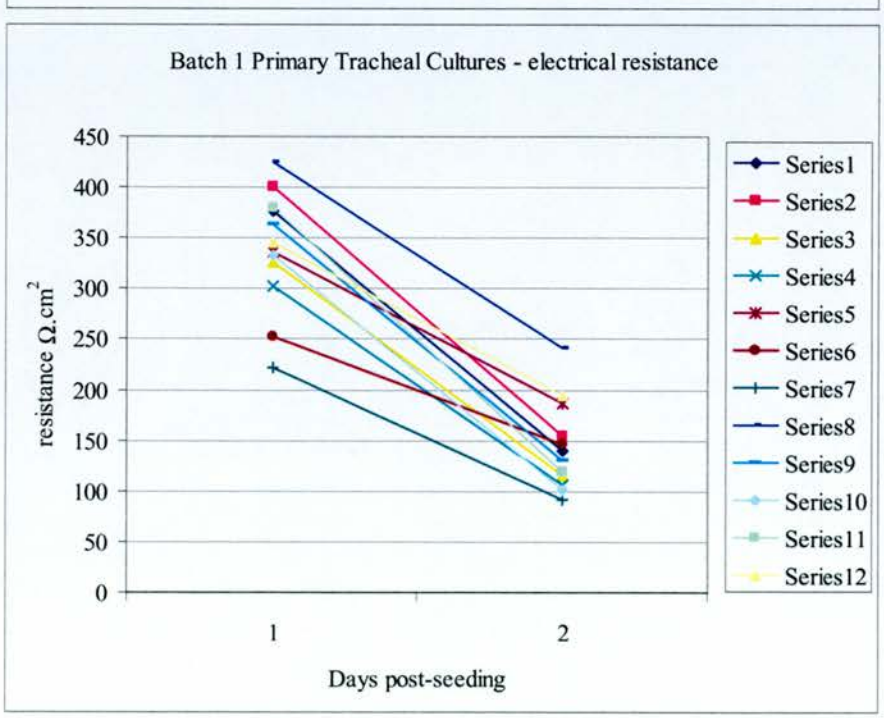
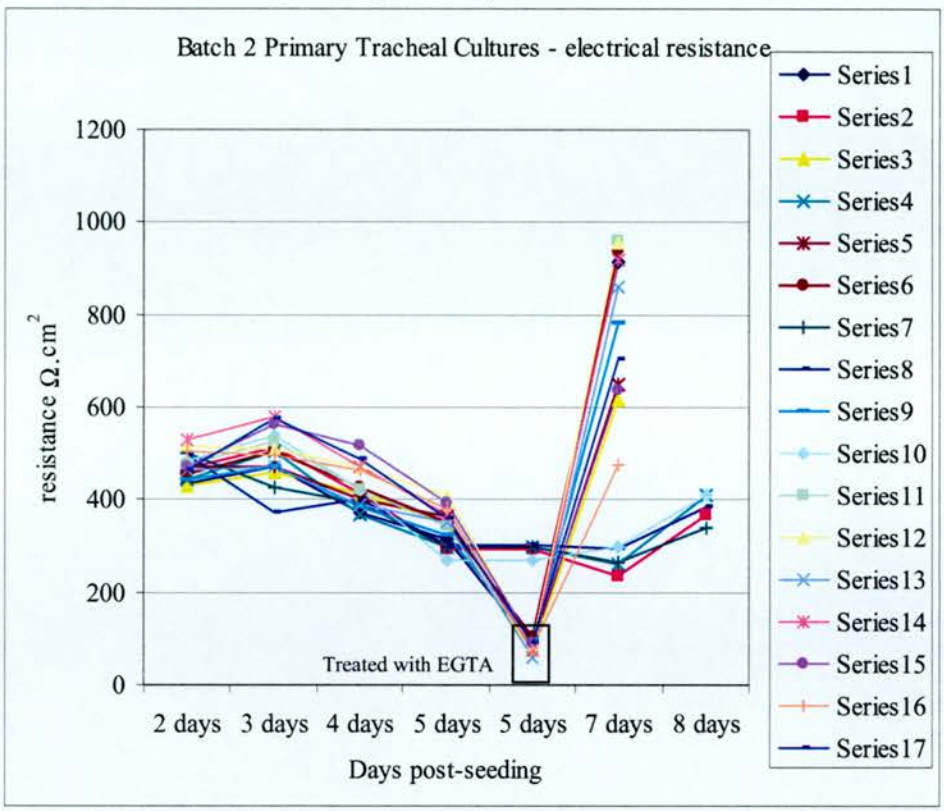


Figure 6.1 (continued)



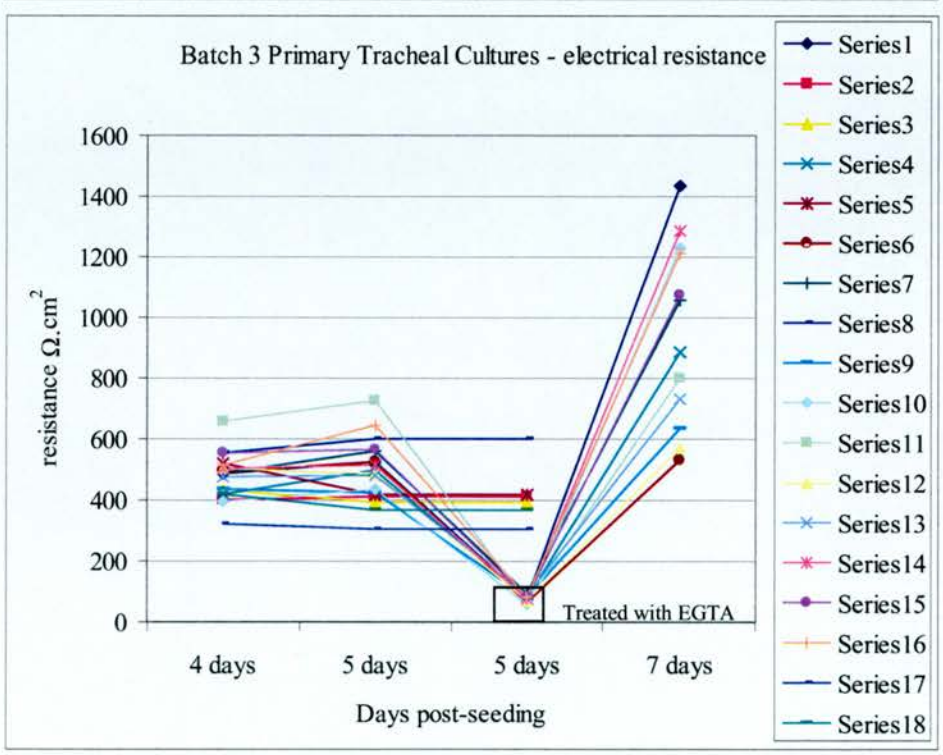
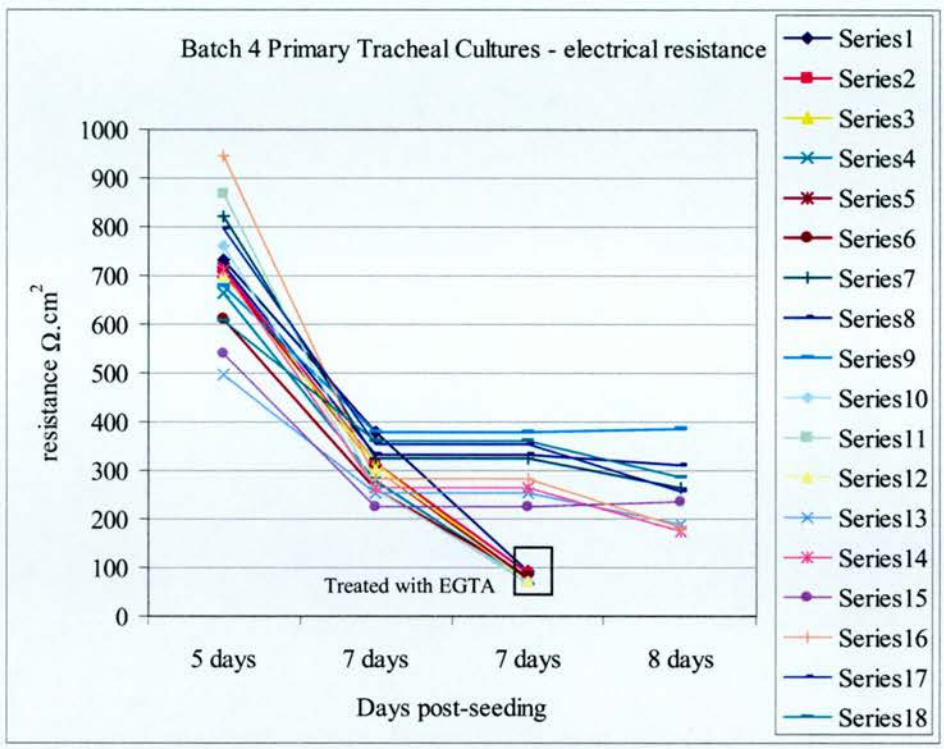


Figure 6.1 (continued)



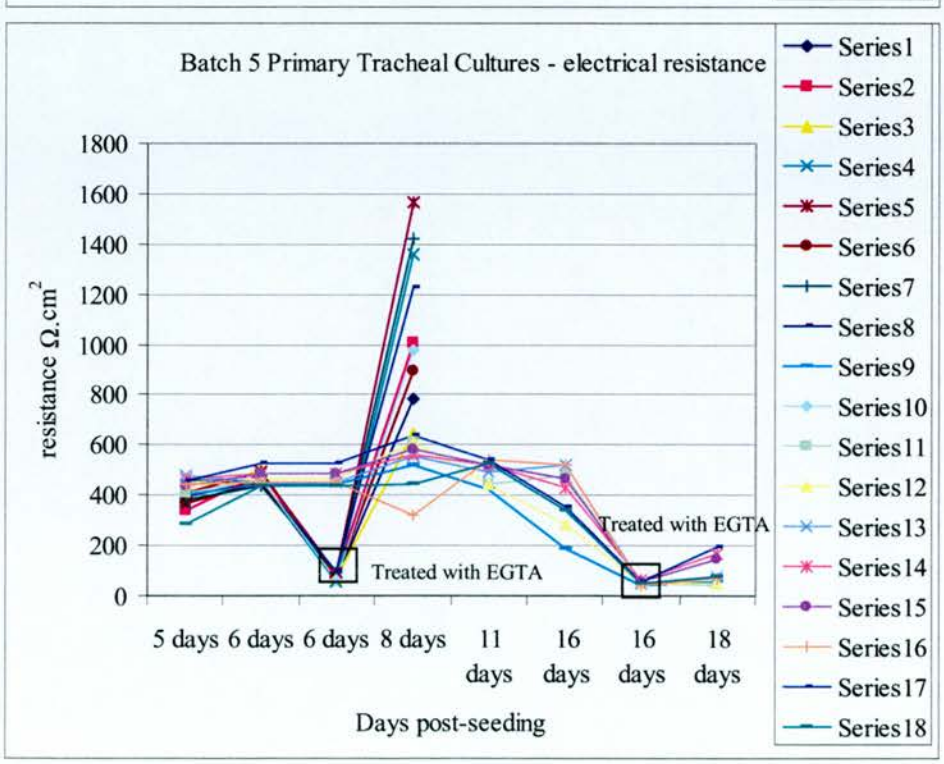
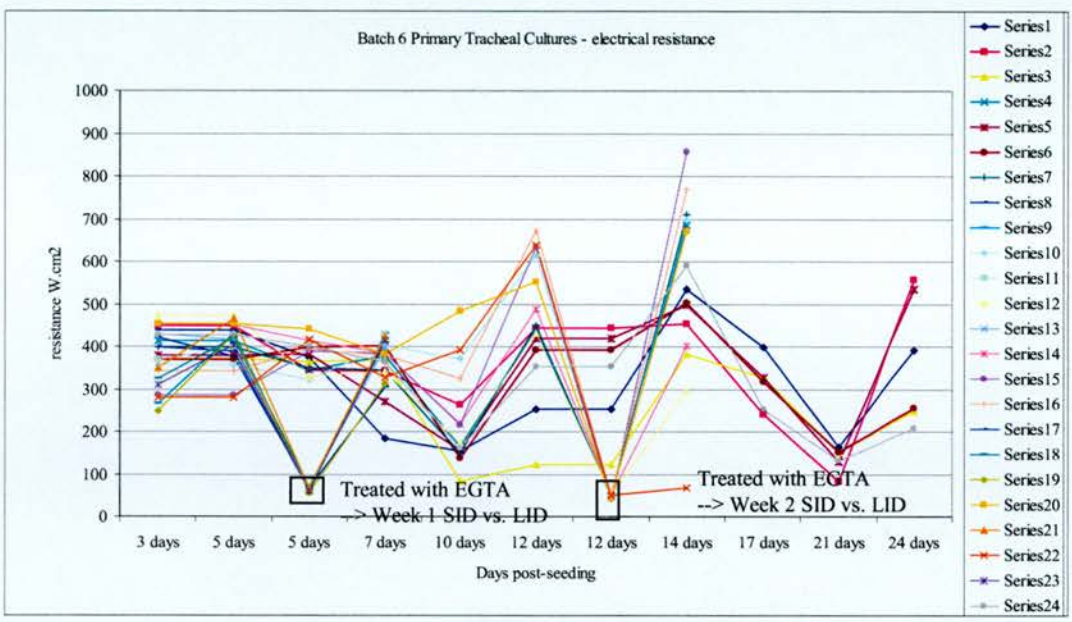


Figure 6.1 (continued)



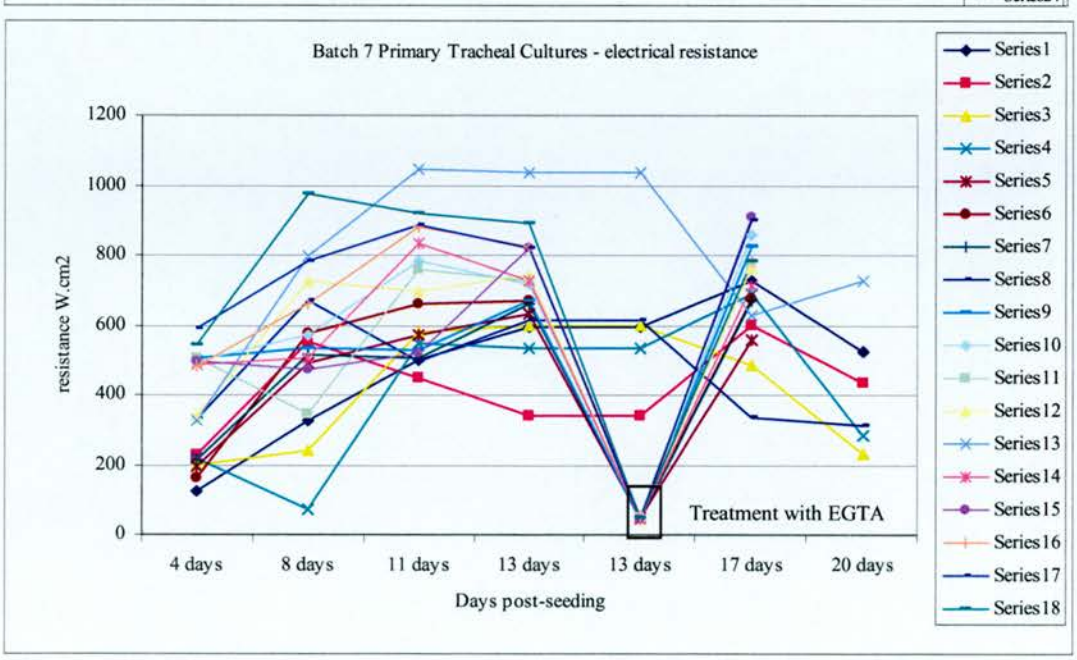
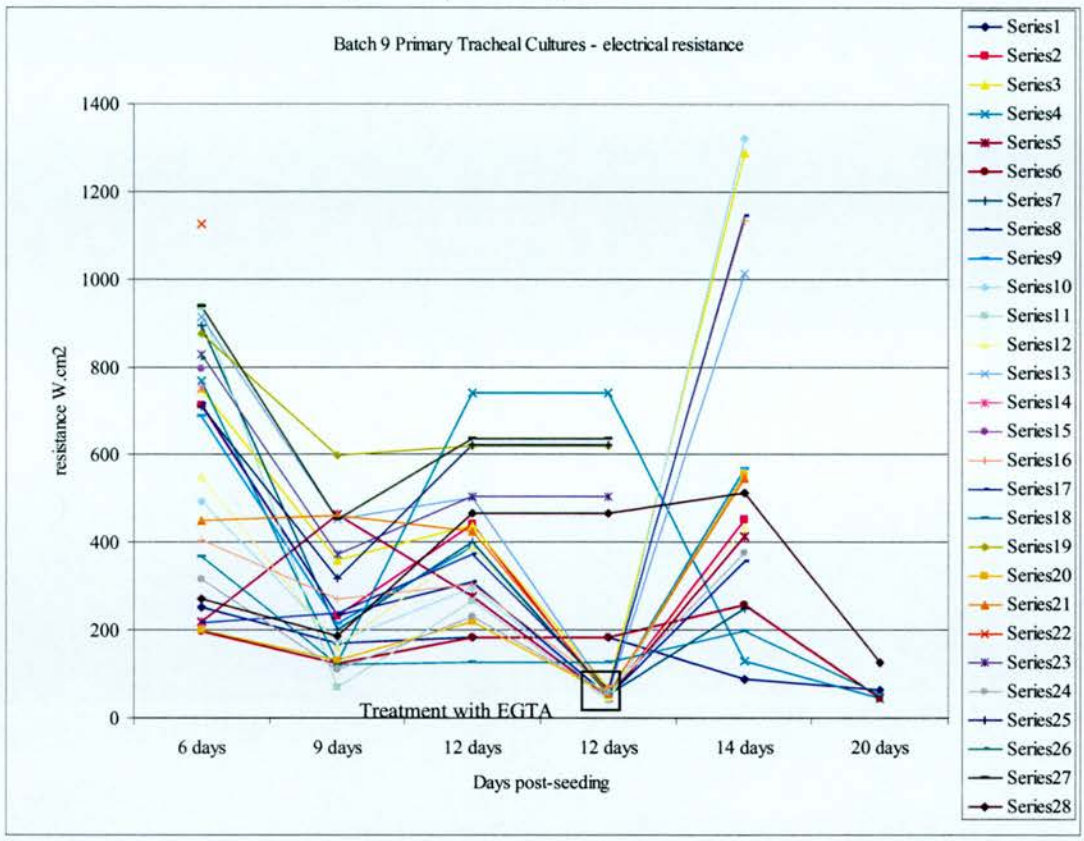


Figure 6.1 (continued)



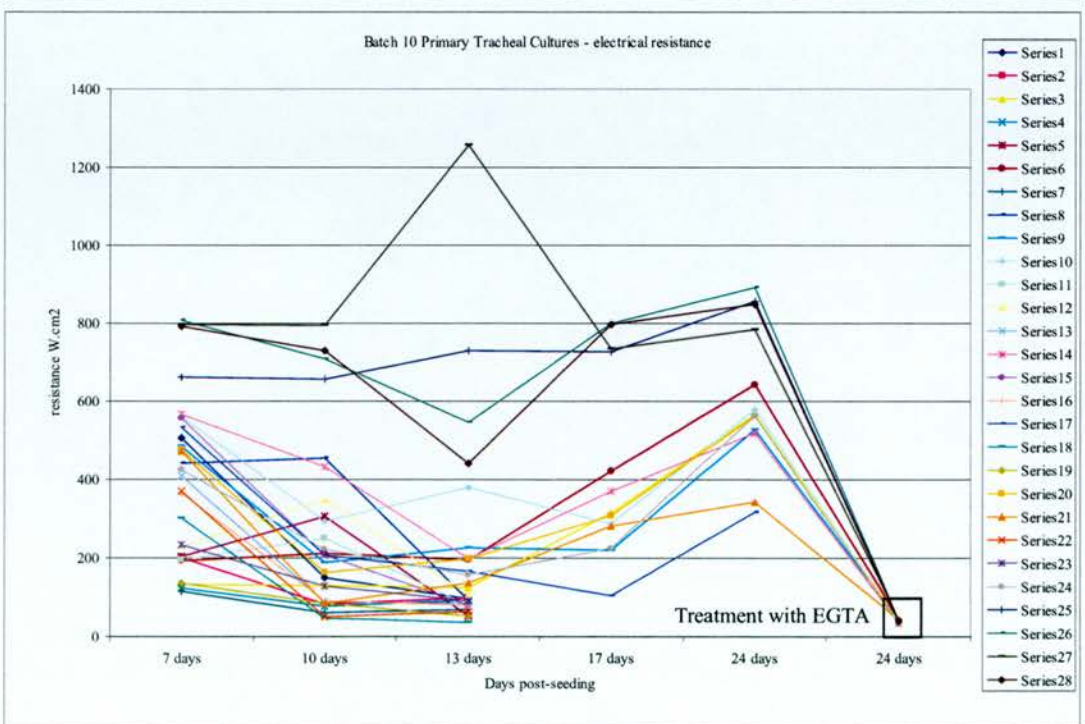
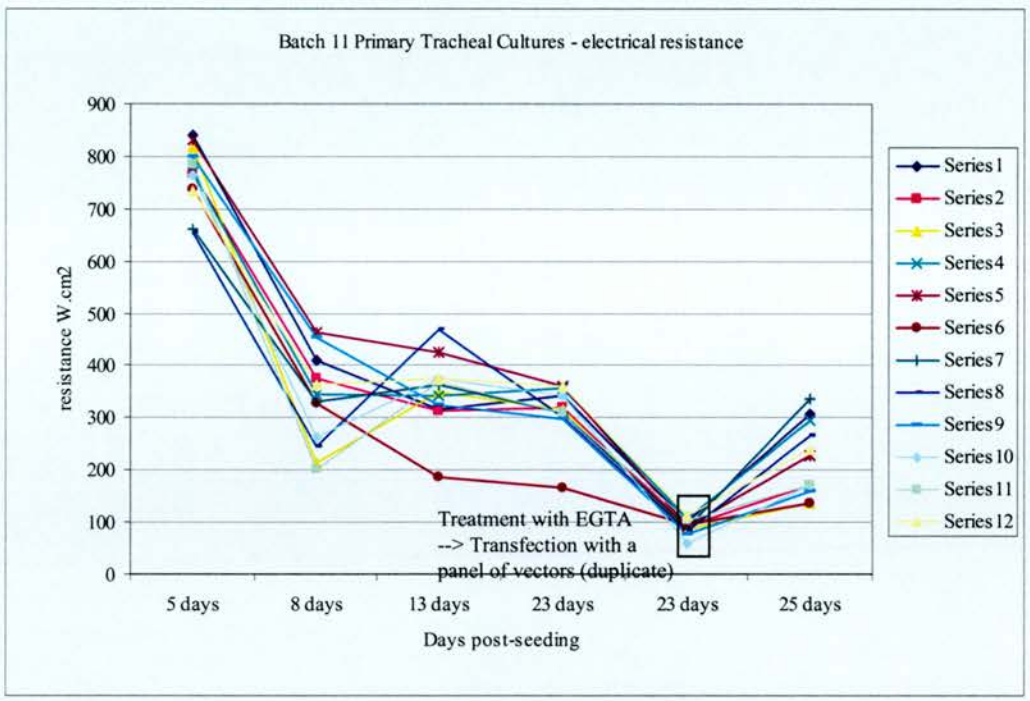


Figure 6.1 (continued)



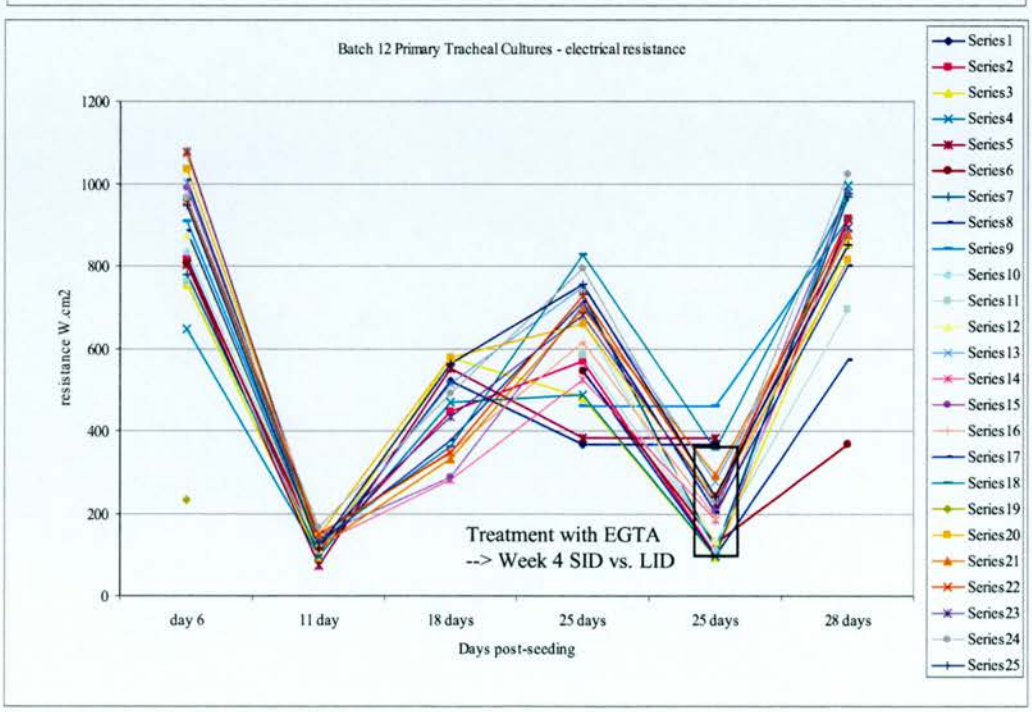
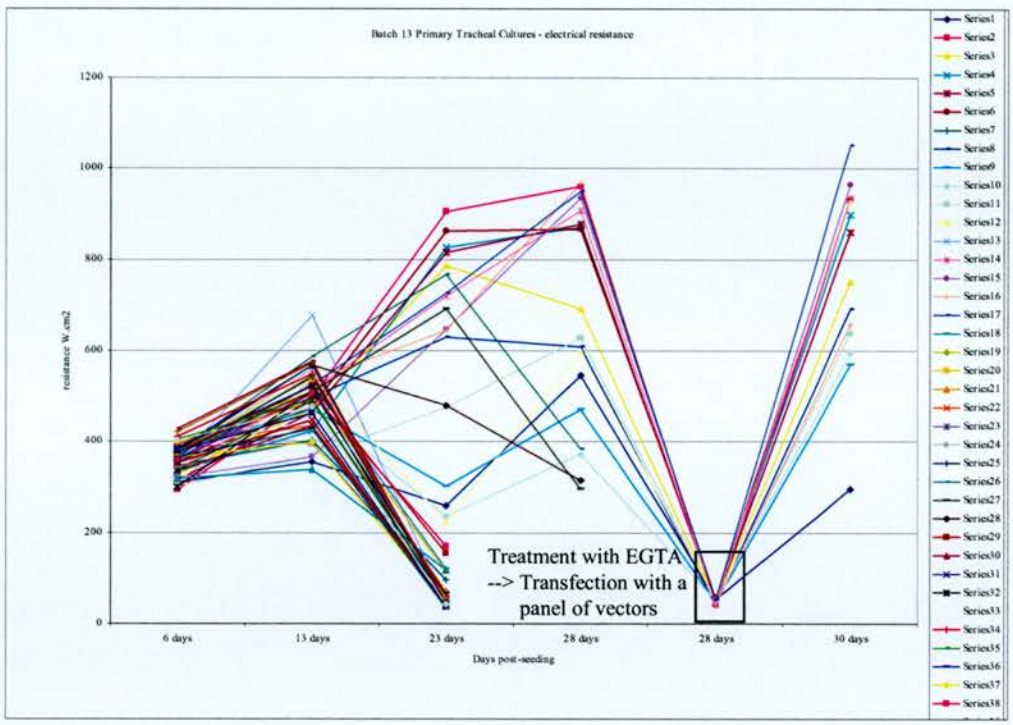


Figure 6.1 (continued)



One might expect EGTA to have a long-term effect reducing the efficiency of tight junction formation and hence resistance. Surprisingly, the opposite was seen. By two days after EGTA treatment, the cells had re-formed tight junctions, and the resistances were often higher than before EGTA disruption. Batch 5 shows a striking example of this; before EGTA treatment, the resistances were between 439 and 526 $\Omega.cm^2$, whereas after recovery from EGTA treatment, the resistances were between 780 and 1566. Cultures from the same batch that were not treated with EGTA did not show a similar rise in resistance at this time point (resistances among these cultures were between 318 and 645 $\Omega.cm^2$). This phenomenon was also observed in batches 2 and 3, and to a lesser degree in batches 6, 7, 9 and 12.

6.2.3 Visual Observations

A number of observations were made when routinely measuring the electrical resistance of the cultures, and changing the growth media, or during routine examinations under the microscope. Most notably, the presence of beating cilia was observed on some cells in almost every culture. Ciliary action was indicated by a shimmering effect over individual cells, or patches of cells, when focusing just above the plane of the epithelial cell sheet under the light microscope.

Detached cells were often seen at the edge of a culture, under the light microscope. These cells were motile, describing regular circular patterns, or exhibiting a jiggling movement. At first, this caused concern over the possibility of contamination with a motile microorganism. However, the moving cells seemed to be of epithelial, rather than microbial, morphology and no other evidence of microbial contamination was present. It seems likely that these were dead epithelial cells that had migrated to the edge of the culture, and were being shunted about by ciliated cells in that region.

Several cultures were ruined when the sheet of cells pulled away from the edge of the culture insert (specifically, from batches 4, 5 and 10). This made it impossible to get an accurate reading of the electrical resistance (these gaps result in a low resistance reading because the electrical current can travel through the gaps at the edge of the culture, rather than transepithelially).

There is some evidence that this was due to insufficient collagen coating of the Costar inserts. In batch 10, almost all of the cultures showed sheets of cells pulling away from the edge of the insert. Cultures # 25 – 28, however, did not show this behavior, and exhibited unusually high resistances. These four cultures had been seeded on inserts that had been collagen-coated at 4°C for several days (as they were left over from a previous batch), while the other cultures were seeded on inserts that had been freshly coated with collagen for only five hours. Thus, it seems likely that insufficient collagen coating resulted in poor cell adhesion to the insert in batch 10. After this observation, the protocol was changed, so that inserts were always collagen-coated at for at least 24-hours before seeding; this seemed to reduce the instances where cultures became detached from the edges.

Fungal contamination was seen in two culture wells (#11 from batch 2 and #19 from batch 12): this appeared a few days after seeding the cultures. The appropriate wells were disposed, and surrounding wells did not become infected. The source of this fungus is unknown, but it is tempting to speculate that it was derived from the sheep animals, as they are not kept in sterile living conditions. The entire batch 8 was discarded as it was found to be contaminated the day after seeding, probably with bacteria (the media had turned yellow).

6.2.4 Electron microscopy

Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) were performed on the cultures for further characterisation. Duplicate samples were selected for both SEM and TEM. Samples were taken at time points of approximately 1, 2, 3, and 4 weeks (the exact dates were dictated in part by the availability of tracheas for seeding and the opportunity for sample submission for electron microscopy processing, thus the time points are not exactly seven days apart). Specifically, 'week 1' is batch 9 cultures #15 + 26 (SEM) and #14 + 23 (TEM) at 6 days post-seeding, 'week 2' is batch 9 cultures #19 + 25 (SEM) and #23 + 27 (TEM) at 13 days post-seeding, 'week 3' is batch 7 cultures #2 + 13 (SEM) and #1 + 8 (TEM) at 21 days post-seeding, and 'week 4' is batch 6 cultures #1 + 5 (SEM) and #2 + 6 (TEM) at 25 days post-seeding.

All samples used for electron microscopy had high resistances (Fig. 6.1), and microscopic examination showed that the morphology was typical for that batch. The initial fix was performed as described in the Methods section. Samples were processed for electron microscopy as a service by Steve Mitchell at the Department of Veterinary Medicine, University of Edinburgh. A parallel analysis of other primary air-interface cultures was done by A. Doherty, and this provided additional confirmation of the findings presented here.

6.2.4.1 Scanning electron microscopy

The apical surface of the cultures was characterised by SEM. Figures 6.2 and 6.3 show scanning electron microscopy images of cultures from weeks 1 through 4.

The cultures predominantly resembled a 'moonscape' of non-ciliated cells at every time point. Very short, fingerlike structures, or bumps, were seen on many cells; these appeared to be microvilli. A few cells exhibited longer hair-like structures, which appeared to be cilia (Fig. 6.2). These ciliated cells were infrequent in the culture, but were present at every time point.

Ciliated cells appeared to be organized into clusters within the 'moonscape'. The number of ciliated cells within each cluster increased as the culture period progressed (Fig. 6.3). At week 1, one to two ciliated cells were seen in close proximity (amounting to approximately 1% of cells in the total culture). At week 2, clusters of three to four cells were seen (approximately 2% of cells in the total culture). At week 3, clusters of six to eight cells were seen (approximately 5 % of cells in the total culture), and by week 4, clusters of eight to twenty-five cells, in nearly confluent patches were observed (amounting to approximately 15% of cells in the total culture).

Duplicate samples were processed, and the abundance of ciliated cells became more variable as the cultures progressed. Thus while the duplicates of week one and two cultures contained almost identical numbers of ciliated cells, there was greater heterogeneity in the number of ciliated cells between duplicates at weeks three and four.

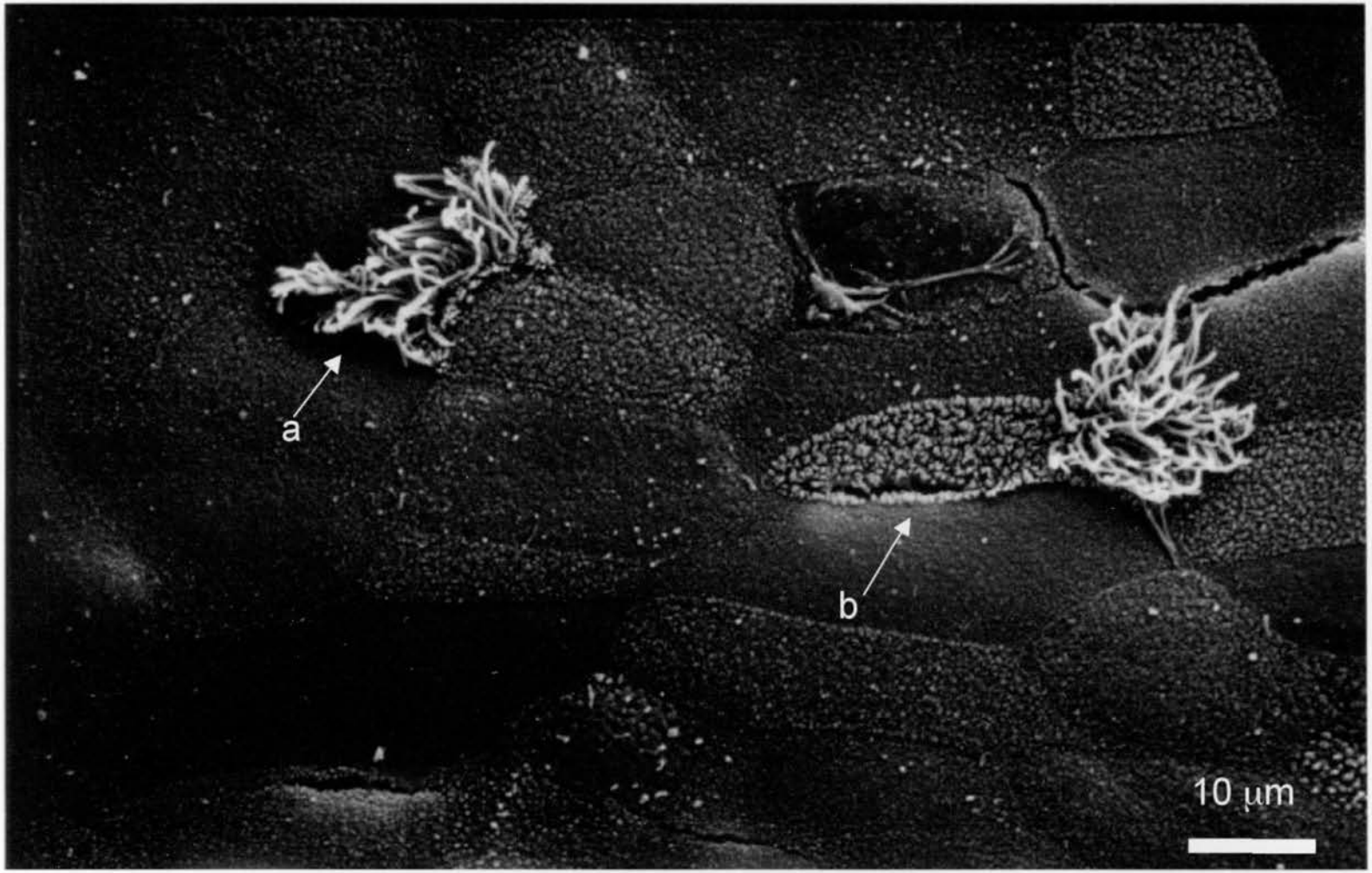


Figure 6.2. SEM of sheep primary air-interface cultures (2.2K magnification). This is a picture of scanning electron microscopy of the apical surface of a week-1 primary air-interface culture. The apical surface resembles a moonscape. **(a)** Ciliated cells and **(b)** Cells with microvilli were present.

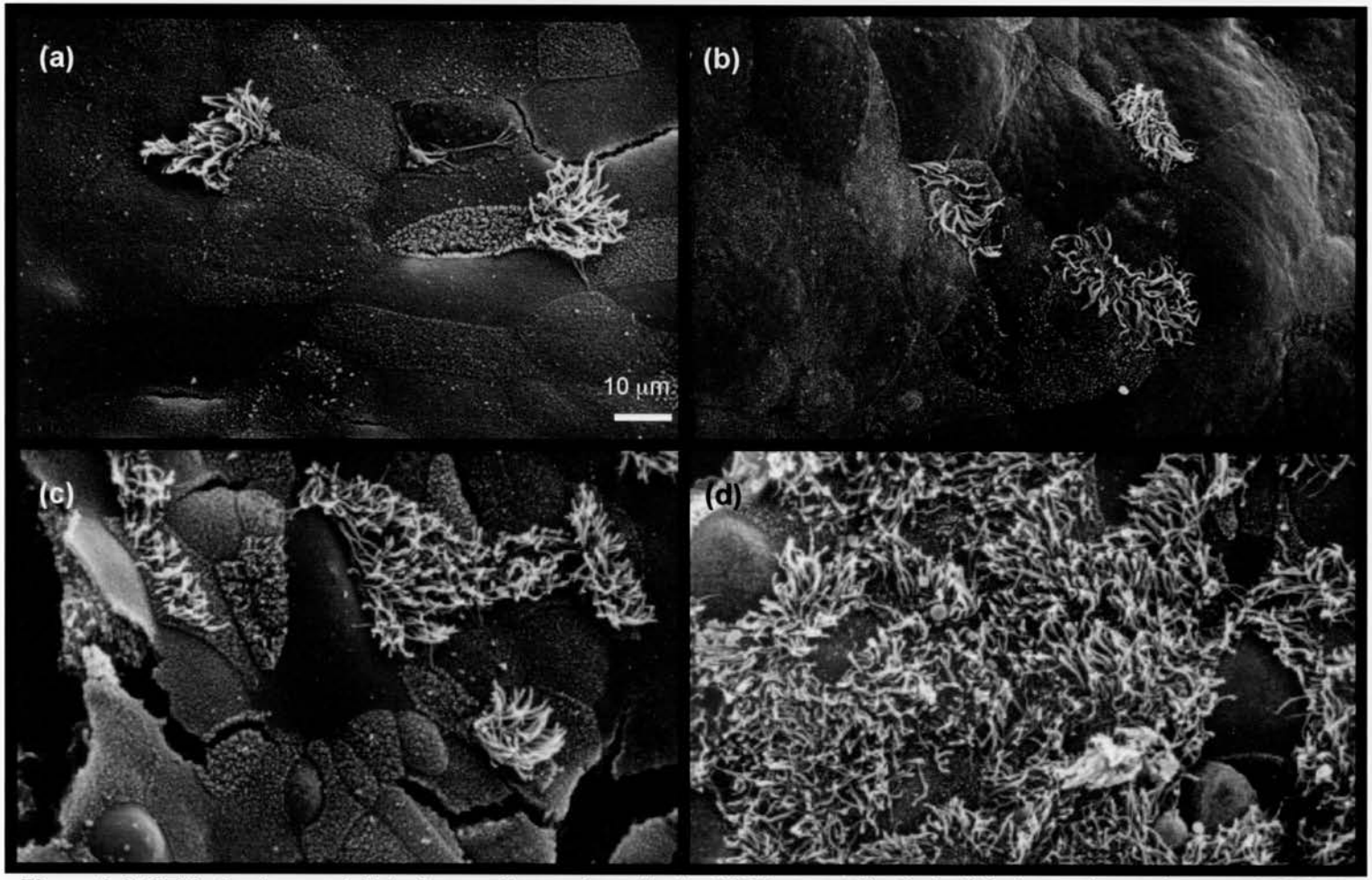


Figure 6.3. SEM of primary air-interface cultures at weeks 1 - 4 (950x magnification). This figure shows pictures of SEM of the apical surface of primary air-interface cultures at: **(a)** week 1 **(b)** week 2 **(c)** week 3 **(d)** week 4. These are not typical regions of the sample, but show patches where ciliated cells were seen (the majority of the culture was non-ciliated at every time point). Ciliated cells became more numerous in the culture from week 1 (patches of 1-2 ciliated cells, amounting to ~1% of cells in the culture) to week 4 (nearly confluent patches of 8-25 cells, amounting to ~15% of cells in the culture).

Hence, these primary tracheal cultures showed features of the endogenous tracheal epithelium on their apical surface, most notably, cilia. Cilia were less abundant than in the native trachea. The week four cultures showed the greatest frequency of ciliated cells, therefore these approximated the best model of the endogenous trachea.

6.2.4.2 Transmission electron microscopy

Figures 6.4 and 6.5 show the transmission electron microscopy images of cultures from weeks 1 through 4.

At every time point, the samples were predominantly composed of a pseudostratified epithelial bilayer (Fig. 6.4). This mimics the structure of the endogenous trachea, where the flatter basal cells lie partially below the columnar cells (see Fig. 1.4). Microvilli could be seen on many cells at the apical surface. These short, finger-like structures appeared to be protrusions of the cell membrane.

Cilia were seen on cells in the week 3 and week 4 samples (Fig. 6.4 and 6.5). These appeared as long, dark structures, embedded in the membrane. These were morphologically distinct from the microvilli; they are darker, wider, and longer, and appear as distinct structures. While the microvilli sometimes appear to branch, the cilia never do. Most convincingly, the 9 + 2 structure of microtubules could be visualised where the sectioning plane cut across the cilia (Fig. 6.4). Cilia were only seen upon a few cells in the specimens examined, although in the week 4 sample, two adjacent cells with cilia were seen.

The appearance of the specimens remained relatively consistent throughout the culture period from one to four weeks. Some patches within the cultures demonstrated additional layers of cells (e.g. in the week 2 culture, Fig. 6.5), or thinner layers with a flatter cells in the lower layer (e.g. in the week 4 culture, Fig. 6.5). However, the majority of the culture was composed of a pseudostratified bilayer at every time point, and no trend was seen as the cultures progressed, relating to a change in cell morphology or arrangement. Although there was some evidence of an increase in the number of ciliated cells from weeks one to week four, it would not be

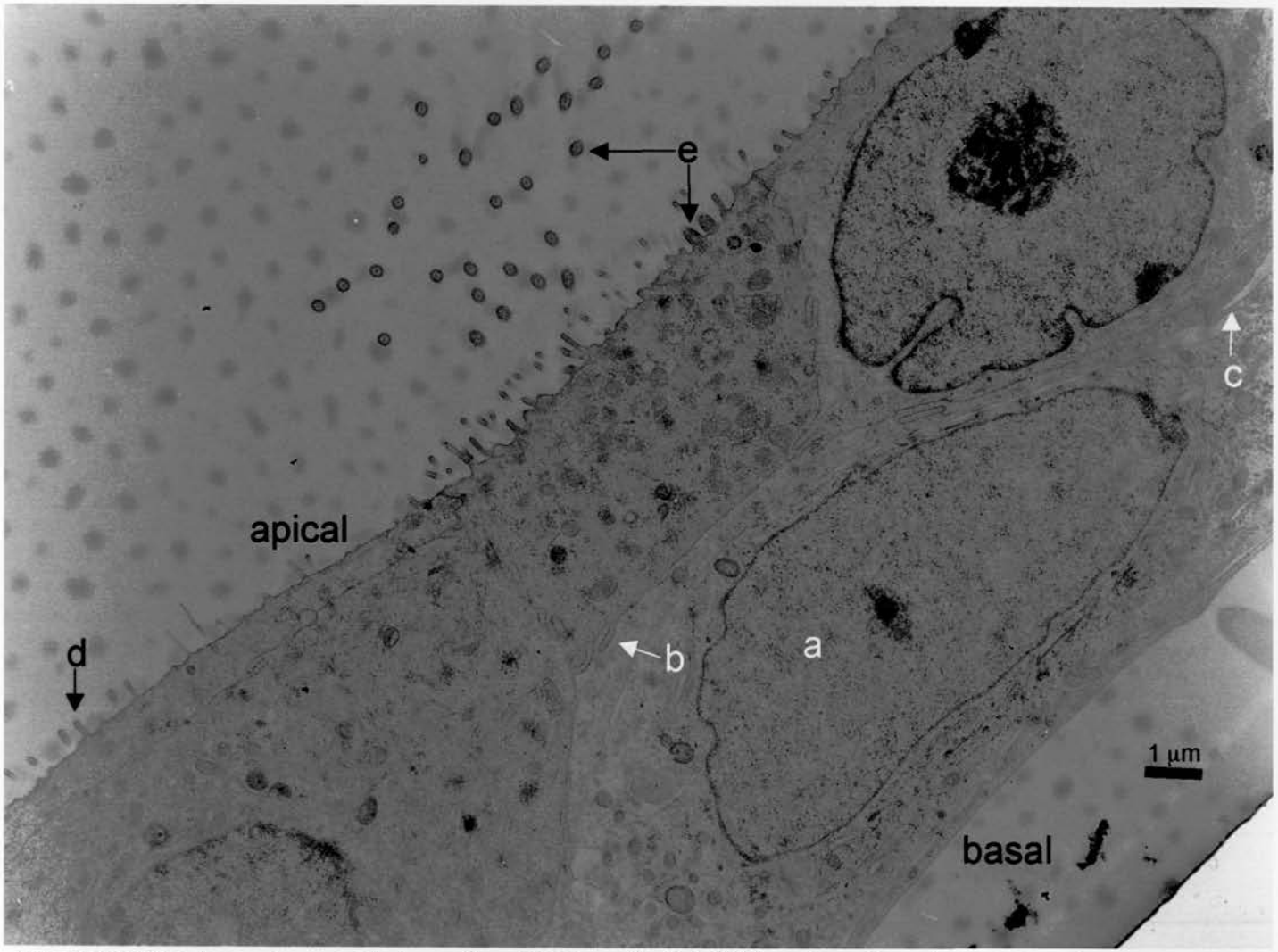


Figure 6.4. TEM of sheep primary air-interface cultures (9.39K magnification). This is a picture of transmission electron microscopy on a week-3 primary air-interface culture. Cells appeared to be arranged in a pseudostratified bilayer (the apical and basal surfaces are labelled). Structures of the native tissue were present: **(a)** nuclei, **(b)** tight junctions, **(c)** gap junctions, **(d)** microvilli and **(e)** cilia (both the root of the cilia in the cell and a cross section of the cilia above the cell can be seen).

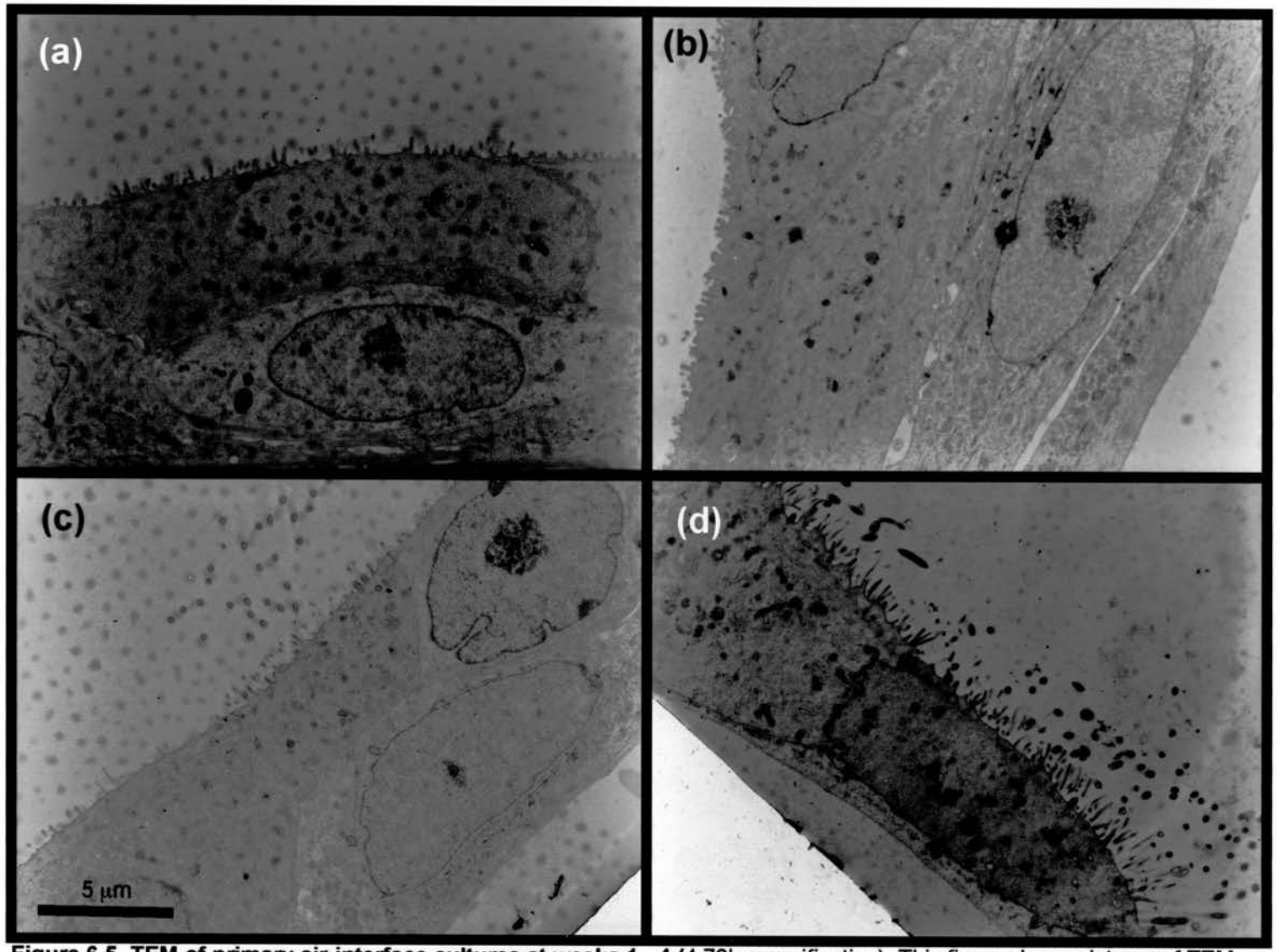


Figure 6.5. TEM of primary air-interface cultures at weeks 1 - 4 (4.76k magnification). This figure shows pictures of TEM of primary air-interface cultures at: **(a)** week 1 **(b)** week 2 **(c)** week 3 **(d)** week 4.

appropriate to use this method for a quantitative measure of ciliated cells, as only a narrow cross-section of the culture was examined for each specimen.

6.2.5 Optimising transfection

Before transfection, the cultures were treated with 6 mM EGTA for twenty minutes to disrupt the tight junctions, as previously described. Two transfection methods, LID and SID, were compared for efficiency in transfecting the air-interface primary cultures. These reagents were compared to roughly optimise transfection efficiency in the primary cultures at time point of week 1, week 2, and week 4:

'Week 1': Nine of the batch 6 cultures displaying high resistances (Fig. 6.1) were treated with EGTA, then transfected with the pEGFP-N plasmid using either the SID or the LID method, at day 5 post-seeding. At day seven post-seeding, FACS was performed.

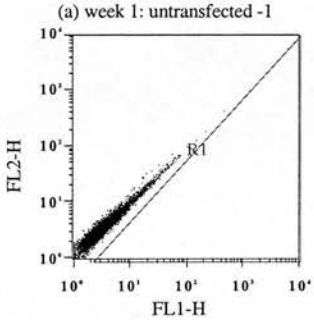
'Week 2': Another nine of the batch 6 cultures displaying high resistances (Fig. 6.1) were treated with EGTA at day 12 post-seeding, then transfected with the pEGFP-N plasmid using either SID or LID At 14 days, FACS analysis was performed.

'Week 4': Nine of the batch 12 cultures displaying high resistance (Fig. 6.1) were treated with EGTA at 25 days post-seeding, then transfected with the pEGFP-N plasmid using either the SID or the LID method. At 27 days post-seeding, FACS was performed.

The result of the FACS analysis for these samples is shown in Figure 6.6. EGFP-expressing cells were detected following both SID and LID transfection of the pEGFP-N plasmid at all three time-points. This project did not allow the extensive period of study required to fully optimise transfection.

Figure 6.6. SID and LID transfection of air-interface primary cultures. Air interface cultures at 1-week, 2-weeks and 4-weeks post-seeding were transfected with the pEGFP-N plasmid using the SID or LID reagent, or left untransfected, as labelled. At 48-hours post-transfection, cells were analysed by FACS (see Methods). This figure shows the dotplots for these transfections.

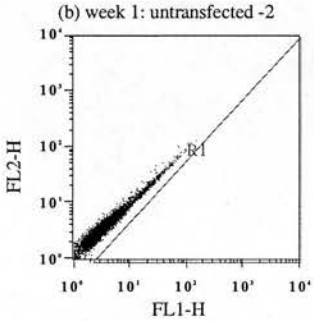
(a) week 1: untransfected -1



Statistics for: 6ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.38
	(X)				4.14
R1	(Y)	0	0.00	0.00	--
	(X)				--

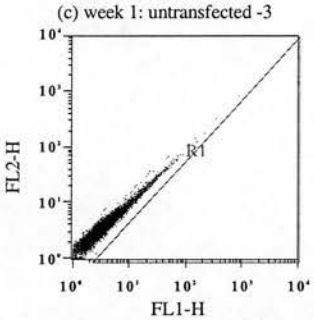
(b) week 1: untransfected -2



Statistics for: 6ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.68
	(X)				4.48
R1	(Y)	0	0.00	0.00	--
	(X)				--

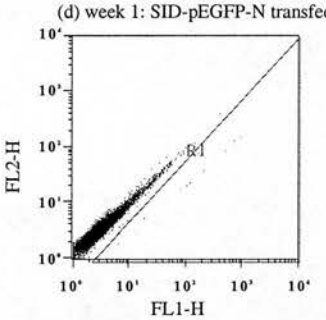
(c) week 1: untransfected -3



Statistics for: 6ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.54
	(X)				4.33
R1	(Y)	0	0.00	0.00	--
	(X)				--

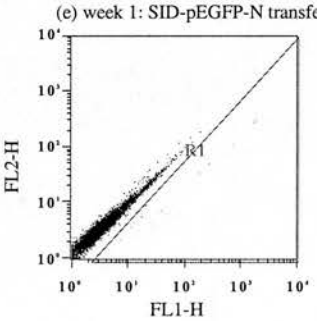
(d) week 1: SID-pEGFP-N transfected -1



Statistics for: 6SNT1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.25
	(X)				4.29
R1	(Y)	12	0.12	0.12	52.25
	(X)				276.20

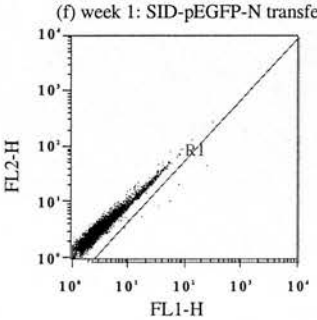
(e) week 1: SID-pEGFP-N transfected -2



Statistics for: 6SNT2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.50
	(X)				4.71
R1	(Y)	14	0.14	0.14	66.81
	(X)				388.77

(f) week 1: SID-pEGFP-N transfected -3

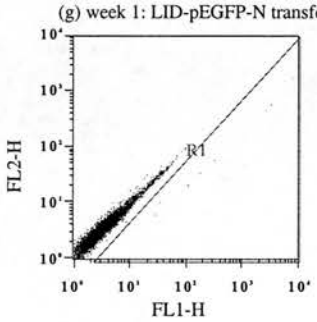


Statistics for: 6SNT3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.32
	(X)				4.29
R1	(Y)	8	0.08	0.08	39.28
	(X)				217.88

Figure 6.6. (continued)

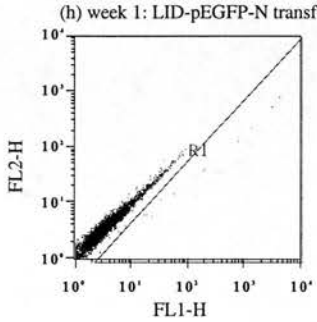
(g) week 1: LID-pEGFP-N transfected -1



Statistics for: 6LID1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	4.33
(X)				5.33
R1 (Y)	9	0.09	0.09	336.66
(X)				1681.50

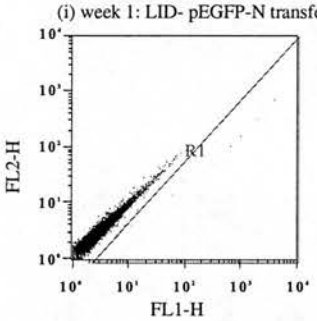
(h) week 1: LID-pEGFP-N transfected -2



Statistics for: 6LID2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	4.82
(X)				6.52
R1 (Y)	15	0.15	0.15	521.69
(X)				1789.81

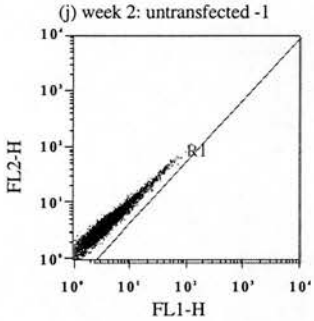
(i) week 1: LID- pEGFP-N transfected -3



Statistics for: 6LID3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	3.74
(X)				4.19
R1 (Y)	8	0.08	0.08	173.79
(X)				979.21

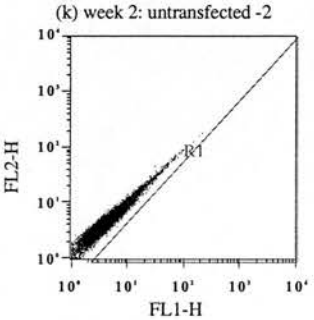
(j) week 2: untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	5.16
(X)				4.76
R1 (Y)	0	0.00	0.00	--
(X)				--

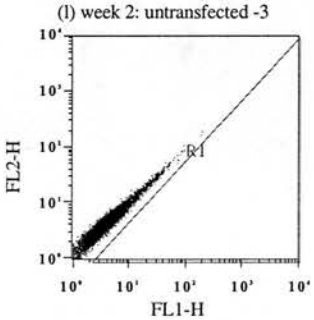
(k) week 2: untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	5.47
(X)				5.12
R1 (Y)	0	0.00	0.00	--
(X)				--

(l) week 2: untransfected -3

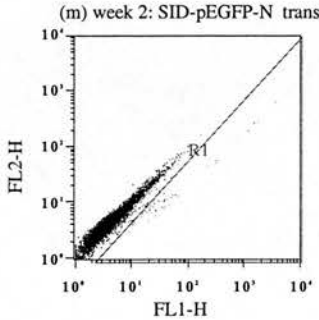


Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	5.20
(X)				4.83
R1 (Y)	0	0.00	0.00	--
(X)				--

Figure 6.6. (continued)

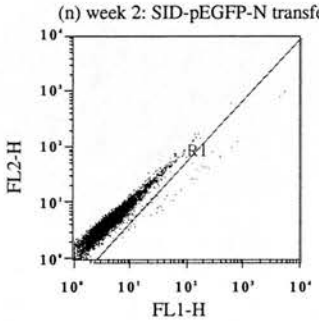
(m) week 2: SID-pEGFP-N transfected -1



Statistics for: SAINTP61 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.33
	(X)				5.90
R1	(Y)	54	0.54	0.54	42.90
	(X)				223.15

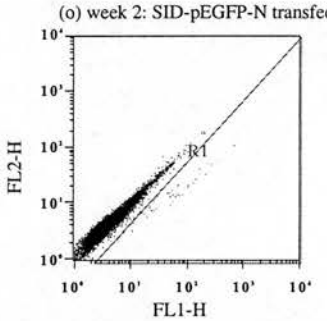
(n) week 2: SID-pEGFP-N transfected -2



Statistics for: SAINTP62 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.75
	(X)				6.62
R1	(Y)	52	0.52	0.52	59.71
	(X)				297.82

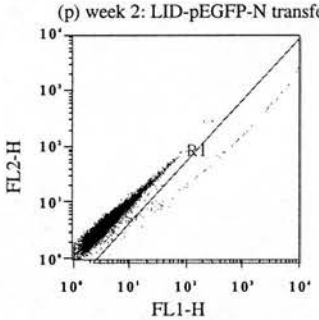
(o) week 2: SID-pEGFP-N transfected -3



Statistics for: SAINTP63 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.41
	(X)				5.35
R1	(Y)	42	0.42	0.42	21.01
	(X)				96.42

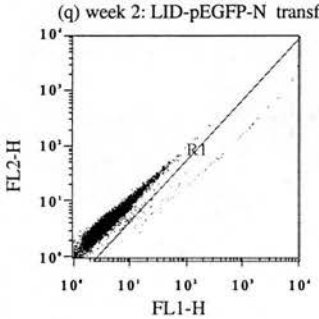
(p) week 2: LID-pEGFP-N transfected -1



Statistics for: LIPOP61 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.14
	(X)				14.79
R1	(Y)	84	0.84	0.84	257.83
	(X)				1217.01

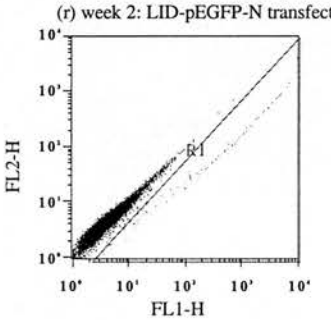
(q) week 2: LID-pEGFP-N transfected -2



Statistics for: LIPOP62 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.40
	(X)				10.80
R1	(Y)	71	0.71	0.71	181.46
	(X)				849.58

(r) week 2: LID-pEGFP-N transfected -3

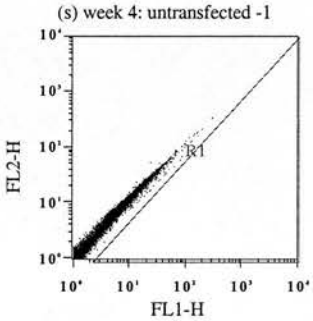


Statistics for: LIPOP63 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.90
	(X)				15.75
R1	(Y)	88	0.88	0.88	295.40
	(X)				1232.80

Figure 6.6. (continued)

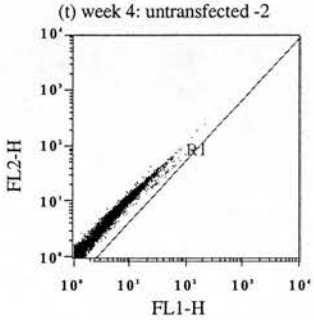
(s) week 4: untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.19
	(X)				7.74
R1	(Y)	0	0.00	0.00	--
	(X)				--

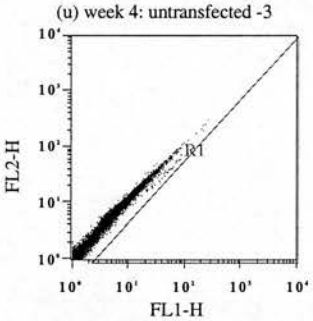
(t) week 4: untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.78
	(X)				6.63
R1	(Y)	0	0.00	0.00	--
	(X)				--

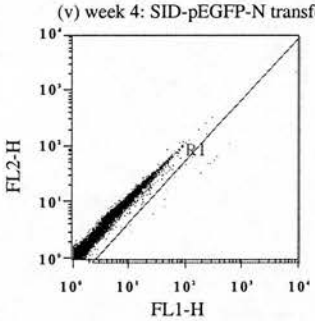
(u) week 4: untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.75
	(X)				7.49
R1	(Y)	0	0.00	0.00	--
	(X)				--

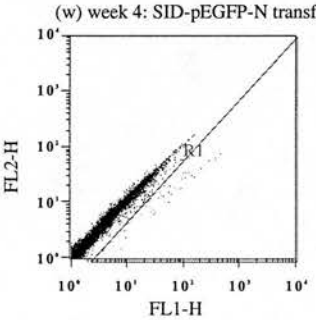
(v) week 4: SID-pEGFP-N transfected -1



Statistics for: SNT1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.14
	(X)				8.57
R1	(Y)	21	0.21	0.21	199.28
	(X)				868.61

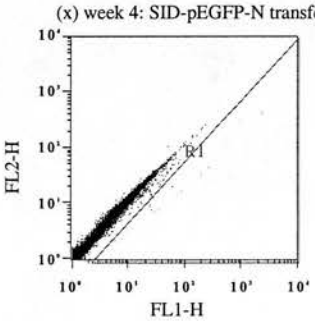
(w) week 4: SID-pEGFP-N transfected -2



Statistics for: SNT2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.11
	(X)				7.42
R1	(Y)	47	0.47	0.47	21.30
	(X)				86.45

(x) week 4: SID-pEGFP-N transfected -3

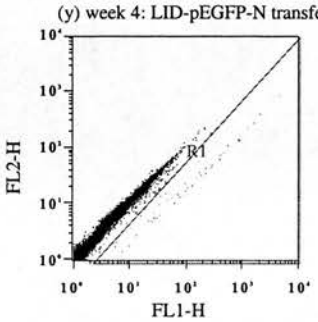


Statistics for: SNT3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.99
	(X)				8.07
R1	(Y)	18	0.18	0.18	57.24
	(X)				253.43

Figure 6.6. (continued)

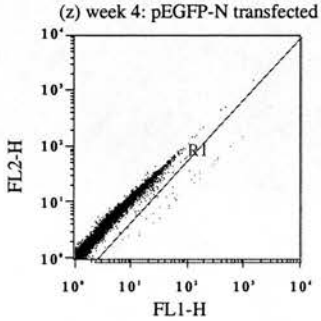
(y) week 4: LID-pEGFP-N transfected -1



Statistics for: EGFPNB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.05
	(X)				9.54
R1	(Y)	59	0.59	0.59	59.78
	(X)				321.78

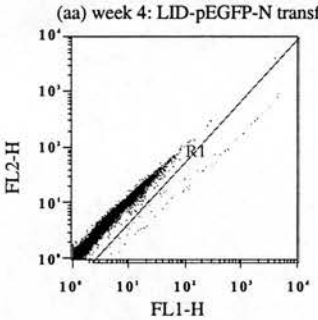
(z) week 4: pEGFP-N transfected -2



Statistics for: EGFPNB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.71
	(X)				8.71
R1	(Y)	50	0.50	0.50	57.66
	(X)				296.77

(aa) week 4: LID-pEGFP-N transfected -3



Statistics for: EGFPNB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.64
	(X)				11.25
R1	(Y)	71	0.71	0.71	93.50
	(X)				494.31

6.2.6 Transfection of primary cultures with a panel of vectors

Electron microscopy studies revealed that week 4 was the time point that most closely resembled the native tracheal epithelium. Thus, this time point was selected for transfection with a panel of genomic context vectors. The LID method was selected for transfection.

Fifteen of the batch 13 cultures showing a high resistance (Fig. 6.1) were treated with EGTA at 28 days post-seeding and transfected with the vectors pEGFP-N, $_{PAC}$RC2cmvEGFP, p1kbcfproEGFP, $_{PAC}$65kbcpfroEGFP, or left untransfected, in triplicate. Cultures with slightly higher or lower resistance were distributed evenly amongst the treatments, to prevent a bias. At day 30, EGFP expression was analysed by FACS (Fig. 6.7). Fluorescent cells were detected following transfection with all four vectors. This experiment was repeated on batch 11 cultures in duplicate at 25 days post-seeding, and a similar result was obtained.

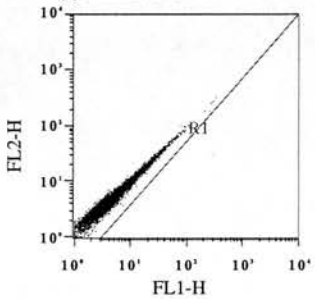
6.3 Discussion

No *ex vivo* model will fully recapitulate the complex interactions of the *in vivo* multi-organ system. Such models must be extensively characterised, to define which features of the native tissue have been retained. However if *ex vivo* cultures are used prudently, bearing in mind their strengths and weaknesses, they can serve a useful function, acting as a bridge between *in vitro* studies on permanent cell lines, which are fairly artificial, and *in vivo* studies, which can be costly and time-intensive. One could take the view that *ex vivo* experiments are superfluous in developing therapeutic treatments, as it will always be necessary to test the efficacy and safety of these treatments in animals before using them in humans. However, from an ethical standpoint it is surely essential to use both *in vitro* and *ex vivo* experiments to reduce, refine and replace the experiments done of animals.

The primary tracheal cultures discussed in this chapter have been characterised to a high degree, as part of this study and other projects. The morphology seen in these cultures is similar to the native tissue; the cells form a pseudostratified bilayer, with evidence of basal and columnar cells. Cilia are present in the culture and become

Figure 6.7. Transfection of EGFP reporter constructs into air-interface primary cultures. Air interface cultures 4-weeks post-seeding were transfected with pEGFP-N, $_{PAC}$RC2cmvEGFP, p1kbcfproEGFP, $_{PAC}$65kbcfproEGFP, complexed with LID, or left untransfected, as labelled. At 48-hours post-transfection, cells were analysed by FACS (see Methods). This figure shows the dotplots for these transfections.

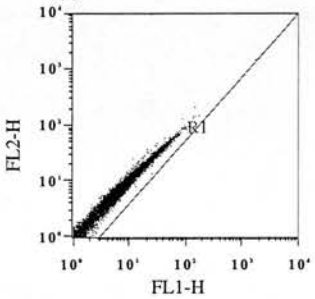
(a) untranfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	6.55
(X)				6.06
R1 (Y)	0	0.00	0.00	--
(X)				--

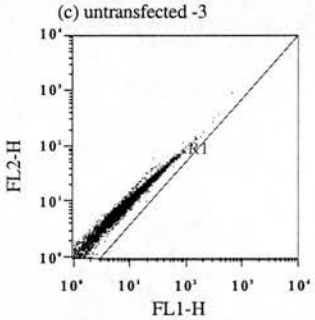
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	10.50
(X)				9.85
R1 (Y)	0	0.00	0.00	--
(X)				--

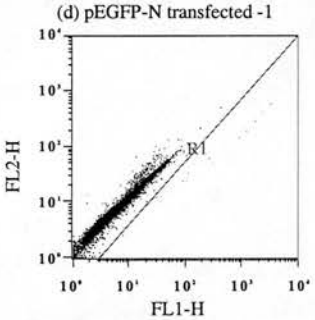
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	11.54
(X)				10.77
R1 (Y)	0	0.00	0.00	--
(X)				--

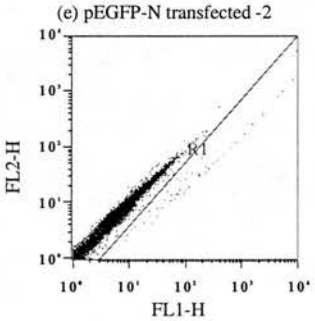
(d) pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	11.12
(X)				10.04
R1 (Y)	17	0.17	0.17	142.31
(X)				788.81

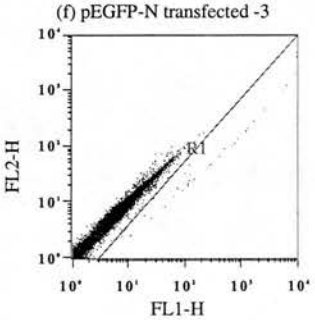
(e) pEGFP-N transfected -2



Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	10.94
(X)				14.79
R1 (Y)	69	0.69	0.69	216.44
(X)				830.36

(f) pEGFP-N transfected -3

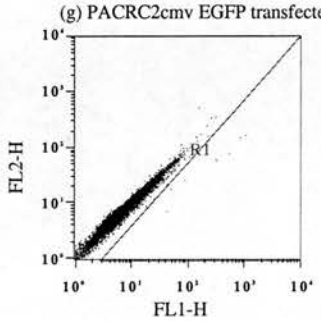


Statistics for: EGFPN4 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	10.19
(X)				13.69
R1 (Y)	39	0.39	0.39	417.51
(X)				1413.20

Figure 6.7. (continued)

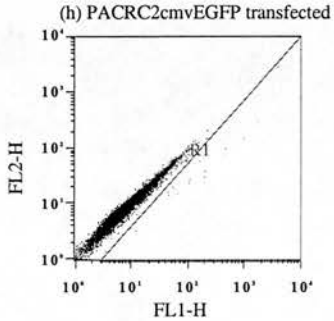
(g) PACRC2cmv EGFP transfected -1



Statistics for: RC2CG1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	12.02
(X)				12.28
R1 (Y)	12	0.12	0.12	889.61
(X)				1611.80

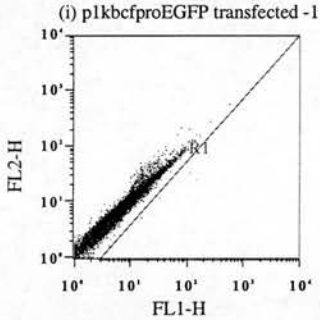
(h) PACRC2cmvEGFP transfected -2



Statistics for: RC2CG2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	10.54
(X)				10.65
R1 (Y)	19	0.19	0.19	227.03
(X)				562.75

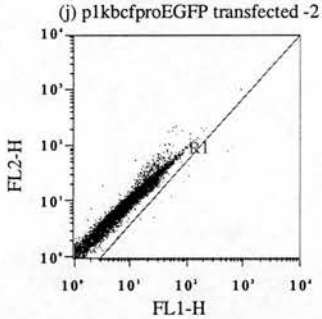
(i) p1kbcfproEGFP transfected -1



Statistics for: 1KB1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	11.23
(X)				10.06
R1 (Y)	5	0.05	0.05	18.64
(X)				71.03

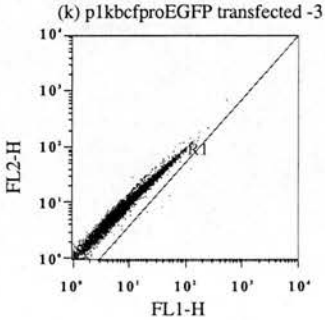
(j) p1kbcfproEGFP transfected -2



Statistics for: 1KB2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	12.18
(X)				11.44
R1 (Y)	12	0.12	0.12	31.30
(X)				150.59

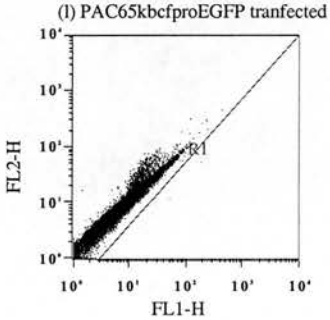
(k) p1kbcfproEGFP transfected -3



Statistics for: 1KB3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	12.28
(X)				11.55
R1 (Y)	7	0.07	0.07	27.75
(X)				136.31

(l) PAC65kbcfproEGFP tranfected -1

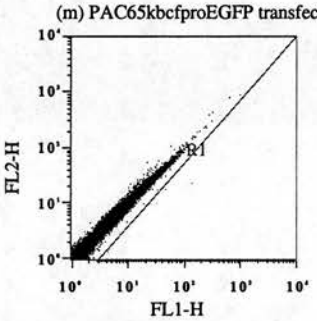


Statistics for: 65KB1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	20000	100.00	100.00	11.86
(X)				10.21
R1 (Y)	2	0.01	0.01	33.22
(X)				119.93

Figure 6.7. (continued)

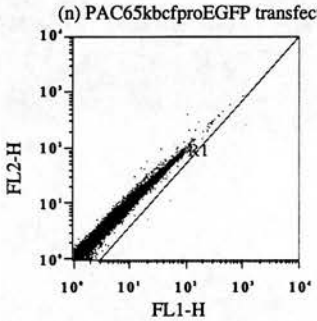
(m) PAC65kbcfproEGFP transfected -2



Statistics for: 65KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	9.90
	(X)				9.19
R1	(Y)	6	0.03	0.03	19.61
	(X)				79.46

(n) PAC65kbcfproEGFP transfected -3



Statistics for: 65KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	11.64
	(X)				10.81
R1	(Y)	3	0.01	0.01	99.55
	(X)				402.42

more numerous with time (although at no time point in this study did they become as numerous as in the native tissue). It was interesting that ciliated cells seemed to develop in clusters: in other words, ciliated cells were usually observed in close proximity to one another (although not necessarily in adjacent cells of direct contact), rather than in a more random scattered arrangement. One might be tempted to speculate that some form of cell to cell signalling is occurring to stimulate the development of these cilia. It would be fascinating to see what the cultures would do given a longer period to develop, or upon the inclusion of secreted cell-signalling factors in the growth media.

A high transepithelial resistance indicated the presence of tight junctions in most of the cultures. A confluent, adherent layer of cells was required to accurately measure the resistance, and a sufficient collagen base was required for this layer to continue to adhere to the insert support. Treatment with 6mM EGTA for 20 minutes was sufficient to temporarily disrupt these tight junctions, although they generally recovered within the following 48-hours.

Successful transfection of the pEGFP-N vector into week 1, 2 and 4 primary cultures was demonstrated, using both the LID and SID reagents.

It would be very interesting to determine the identity of the transfected cells, one preliminary attempt was made to do this (by separating the transfected population according to FL1-H: FL2-H ratio, with a FACSCalibur machine, then adhering the cells to a slide, and staining with Hematoxylin/Eosin to visualise cell structures such as cilia), sadly this was unsuccessful.

Week four cultures were successfully transfected with a panel of EGFP-expressing vectors. There was evidence of EGFP expression with the pEGFP-N, $_{PAC}$RC2cmvEGFP, p1kbcfproEGFP and $_{PAC}$65kbcfproEGFP vectors.

It was unfortunate that these studies were completed during the 2001 period when the foot and mouth epidemic was limiting animal movement and availability in Britain. Indeed, this halted the *ex vivo* phase of the project for a couple of months, and limited the number of sheep tracheae available for use in this study.

Chapter 7: Conclusions and Future Work

7.1 Conclusion

7.1.1 Vector construction (Chapter 3)

Two small plasmids (p1kbcfproEGFP and pCFTRiresEGFP) and four large PAC vectors ($_{PAC}$65kbcfpro, $_{PAC}$RC2cmvEGFP, $_{PAC}$RC1iresEGFP, and $_{PAC}$RC2iresEGFP) were created, coupling CFTR sequences to the EGFP reporter gene. These vectors were used in transfection studies, to evaluate the effects of CFTR regulatory elements upon transcription, and to analyse the suitability of EGFP as a reporter gene for CFTR promoter activity.

The small plasmids were made by conventional cloning procedures, while the large PAC vectors were made by a double-recombination method incorporating both homologous recombination and Cre/loxP recombination, as previously described (Boyd et al, 1999b). The greatest number of recombinant clones was derived by using a high insert: vector ratio in this double-recombination method.

Two aspects of vector construction have been validated by this project: (1) the successful creation of $_{PAC}$65kbcfproEGFP, $_{PAC}$RC2cmvEGFP, $_{PAC}$RC1iresEGFP, and $_{PAC}$RC2iresEGFP showed that it is possible to use a PCR product as an insert in the double recombination method; (2) there was originally doubt as to whether the loxP sequence could be successfully introduced onto the end of a PCR product by its inclusion in the a primer: one might predict that the palindromic loxP site would anneal to itself to form a loop during the PCR process and would thus be deleted from the product in subsequent cycles (see Section 3.2). The successful generation of these PAC vectors shows that it is possible to introduce this sequence by its inclusion in a primer, despite the palindromic nature of the loxP sequence. These innovations in vector construction should prove useful in the creation of further vectors.

Further to this, colony PCR was shown to be a suitable method for detecting recombinant clones generated by the double recombination method. It proved possible to perform PCR-sequencing directly from the product of a colony PCR. These methods will allow rapid and efficient detection and analysis of recombinant clones in the future.

Where a PCR-step was included in the construction process, the corresponding region was sequenced to check for mutations. Despite the use of Expand™ high fidelity Taq polymerase and HPLC-purified primers, many mutations were detected. The vectors reported here were found to be mutation-free with the following exceptions: (1) every $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP clone that was sequenced contained the mutation G → A in the IRES portion of the vector. *ApaLI* restriction has demonstrated that this mutation appears to derive from the parent vector. Thus, the published sequence for the pIRES2EGFP vector is probably incorrect at this basepair; (2) the $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP vectors both contain the mutation ΔT (at positions 72127 and 96232, respectively), between the polyadenylation site of EGFP and the loxP site. This region is derived from the loxP primer; hence, his mutation may reflect aberrant synthesis of the loxP primer, such that a proportion of the primers contain the divergent sequence. This mutation does not affect transcription, and is therefore inconsequential.

7.1.2 Optimisation of transfection efficiency (Chapter 4)

Several reagents were compared to optimise transfection efficiency in the COS-7, MDCK-IOWA, T84, Caco-2, and HBE cell lines. Initial experiments used a βgalactosidase plasmid, with an ELISA or luminometer assay to detect expression. In subsequent experiments, cells were transfected with EGFP-expressing plasmids and expression was measured by FACS. FACS has proved to be a superior method for analysing expression as this method allows the dissection of the number of cells expressing versus the level of expression per cell.

This report demonstrated that EGFP expression could be measured by a cell's FL1-H: FL2-H ratio. In cells that do not express EGFP, the fluorescence ratio is approximately 1:1. However, following transfection with an EGFP-expressing plasmid, an additional population of cells appears where FLH-1>FLH-2. RT-PCR showed that this population represents cells expressing EGFP.

The 1:1 ratio of FL1-H: FL2-H in untransfected cells suggests that where cellular components produce background fluorescence, they do so evenly across the spectrum. Thus, a cell with low background fluorescence has both low FL1-H and

FL2-H while a cell with high background fluorescence has both high FL1-H and high FL2-H. In contrast, the Green Fluorescent Protein does not emit light evenly across the spectrum, but rather has a peak of emission at ~488nm, in the green region of the spectrum, hence making a greater contribution to a cell's FL1-H fluorescence than FL2-H fluorescence. Thus in cells expressing EGFP, FL1-H>FL2-H.

Based on this relationship, a dual-dotplot method of analysis was developed to analyse the FACS data. This method proved to be superior to the traditional histogram analysis, as it allowed a more robust distinction of EGFP-expressing cells (including those expressing low levels of EGFP). With this dotplot method, the difference between the transfected and untransfected populations is qualitative, rather than quantitative. Thus, a single transfected cell can be robustly detected in a sample of 10,000 cells or greater. Semi-quantitative RT-PCR showed that the FL1-H intensity of the cells is probably a direct measure of the level of EGFP expression in these cells, in keeping with the report by Subramanian and Srienc, 1996.

The Saint-Mix™ transfection reagent showed superior transfection efficiency, in comparison to the DOTAP liposome. A ratio of 20 µl Saint-Mix™: 1 µg DNA was found to produce the best results. Although the Saint-Mix™ reagent showed promise at the start of the study, this reagent produced inconsistent results, and the transfection efficiency generally declined throughout the course of the study; the reasons for this were unknown.

In initial experiments utilising the pCMVβ plasmid, the PEI reagent appeared to be superior to the Saint-Mix™ reagent, producing more βgalactosidase in the lysate of transfected cells. However, contradictory results were found when EGFP-expressing plasmids were transfected, and expression was analysed by FACS. In this case, Saint-Mix™ transfected a larger proportion of cells than PEI did. It appears that in the presence of a strong P_{CMV} promoter, PEI might artificially boost expression, as the average fluorescence intensity of the transfected population was high higher with PEI than with Saint-Mix™, in pEGFP-N transfections. The same did not apply to p1kbcfproEGFP transfections.

A new transfection method, the SID method, was developed for this work. SID combines the Saint-Mix™ reagent with the integrin-targeting peptide P6. The inclusion of the P6 peptide increased the efficiency of transfection above that observed with Saint-Mix™ alone, the optimum ratio for transfection was found to be 20 µl Saint-Mix™: 1 µg P6: 1 µg DNA, for most cell lines. This method produced efficient transfection in many cell lines, but sadly showed the same inconsistencies reported for the Saint-Mix™ method. Thus while SID shows great potential, the causes of this inconsistency would have to be pinpointed and eliminated before this method would be suitable for routine transfection studies.

The LID method (combining 0.75 µl of the liposome Lipofectin with 4 µg of the integrin targeting peptide P6, and 1 µg of DNA) was found transfect the greatest percentage of cells in every cell line except HBE, and produced consistent results. This method was selected for further transfection studies.

7.1.3 In vitro *transfection (Chapter 5)*

The reporter constructs described in Chapter 3 were transfected into cell lines to analyse the effects of CFTR regulatory elements upon EGFP expression.

The 797 bp region 5' of the CFTR gene in the p1kbcfproEGFP vector drove a low level of EGFP expression in every cell line tested. This is consistent with reports in the literature, demonstrating that this region encompasses the proximal CFTR promoter (Yoshimura et al 1991b). This study shows that the proximal CFTR 5' region maintains its function in a transient transfection system, and could be used as a promoter in gene therapy vectors.

There was no evidence for modulation of expression by the additional upstream region in the $_{PAC}$65kbcfproEGFP vector. However, it is not possible to rule out a subtle modulation of expression (such as that reported by Nuthall et al, 1999b), due to the complications of vector size effects in this study.

The intron 1 DHS did not appear to affect expression in the COS-7 or MDCK-IOWA cell lines. An analysis of the effects of this element in gut cell lines was frustrated by poor transfection efficiency of those lines, coupled with reduced expression from

large PAC vectors, a weak CFTR promoter, and inefficient translation from an IRES. In previous reports, this element has been shown to increase expression in a transient transfection system, in the Caco-2 and CHO-K1 gut cell lines, but not in a bronchial cell line that expressed CFTR at high levels (Smith, AN et al, 1996; Mogayzel and Ashlock, 2000). In addition, analysis of DNase 1 hypersensitivity shows this element to be specific to cells of the gut, in contrast to the lung (Smith, A et al, 1996). It is particularly interesting that the intron 1 element did not upregulate expression in the MDCK-IOWA kidney cell line. This provides further evidence that the intron 1 is not ubiquitous, and does not upregulate expression in the kidney, despite the fact that this tissue expresses high levels of CFTR.

Large PAC vectors exhibited reduced expression in comparison to small plasmid vectors in permanent cell lines, as shown by a comparison of the vectors $_{PAC}$RC2cmvEGFP and pEGFP-N. Both the number of cells expressing EGFP and the average fluorescence intensity of this population were reduced with the PAC vector, in comparison to the small plasmid vector. This is not an effect of the reduced gene copy number in PAC transfections, and does not affect co-transfected plasmids, as demonstrated by mixing experiments.

Mixing experiments revealed an interesting phenomenon: diluting a reporter plasmid with an anonymous DNA molecule actually *increased* expression levels. This was true for both PAC and plasmid reporter constructs. The anonymous DNA molecule was essential, and had to be mixed with the reporter construct before complex formation for this increase in tranfection efficiency. While we cannot yet explain the mechanism of this effect, there are many potential applications.

7.1.4 Characterisation and transfection of primary cultures

This report showed that that the sheep primary tracheal air-interface culture system demonstrates features of the native tracheal epithelium and is a reasonable model system for gene therapy studies. Electron microscopy demonstrated a pseudostratified epithelial-cell arrangement, similar to the native airway surface epithelium. Some cells in the culture exhibited cilia; these became more numerous as

the culture progressed. Tight junctions were demonstrated in the cultures, by virtue of high transepithelial resistance.

Both SID and LID transfection methods were successful in transfecting the primary cultures. The LID method produced a greater level of expression than the SID method in these experiments, hence this method was selected to transfect a panel of vectors into then primary cultures. The pEGFP-N, p1kbcfproEGFP, $_{PAC}$RC2cmvEGFP and $_{PAC}$65kbcfproEGFP vectors all demonstrated expression when transfected into week 4 primary cultures, using the LID method. A comparison of transfection efficiencies demonstrated that while PAC vectors show greatly reduced expression in comparison to their small plasmid counterparts in permanent cell lines; the difference is not as severe in *ex vivo* cultures. Thus, PAC GCVs may be suitable candidates for *in vivo* trials.

EGFP has proved to be a valuable reporter gene in the analysis of CFTR promoter elements. The ability to distinguish between the percentage of cells expressing EGFP and the level of expression per cell, and the greater sensitivity of detection afforded by FACS analysis were crucial to these studies.

7.2 Future work

7.2.1 EGFP detection

The FACS dual-dotplot analysis method proved to be a superior detection method. Further refinements to the method may increase the sensitivity of detection even further. For example, it could be possible to create a 3-dimensional plot of all three fluorescence readings taken by the FACS machine (FL1-H, FL2-H and FL3-H), this may allow an even more specific dissection of the transfected population. Here, the fluorescence settings on the FACS machine were at 438 for all fluorescent channels (see Methods for machine settings). It appeared to be necessary to keep the FL1-H and FL2-H settings equal, and set the machine to 0% compensation (corrects for fluorescence from alternate channels) to maintain a 1:1 ratio of these fluorescence readings (this is not surprising). However, it may be possible to get a better dissection of the transfected population by altering these settings.

Fluorescence microscopy was attempted early in the project. Transfection of reporter constructs such as pEGFP-N (where a large number of cells expressed high levels of EGFP) produced fluorescent cells that could be visualised under the microscope. However, it was difficult to distinguish fluorescent cells when the PAC constructs were transfected, due to the small numbers of cells expressing EGFP and the low fluorescence intensity in these cells. In addition, fluorescence microscopy is primarily qualitative, rather than quantitative. Hence, this method proved unsuitable for this work. However, it may be possible to develop software for the digital fluorescence microscope to improve this method for our purposes. For example, it may be possible to compare the ratio of fluorescence from different channels, as was done in the FACS dotplot, and perform subtraction analysis based on this ratio to correct for background fluorescence. Applying such software to the digital image might create a visual picture of green fluorescence that is more representative of EGFP expression.

7.2.2 Vector construction

The $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP vectors were created to test the effects of the intron 1 DHS upon expression. This approach was selected for a number of reasons. We hypothesized that CFTR may require its endogenous location, or may require the presence of other genomic elements, to accomplish regulation of expression. Hence, it seemed most appropriate at the start of the project to preserve this intron in its natural context, by maintaining as much of the intact CFTR locus as possible. It proved to be easy to create such constructs by altering the $_{PAC}$RC1b and $_{PAC}$RC2b vectors with the double recombination method.

Before the construction of these vectors, Heather Davidson (a member of the Edinburgh MMC group) created the vector $_{PAC}$RC2EGFP. This vector was engineered to encode a fusion protein, containing CFTR exon 1, intron1 and a portion of exon 2, fused at the N-terminus to the EGFP gene; with 65 kb of CFTR 5' region driving expression (this vector was created using the double-recombination method). This vector was transfected into the COS-7 and MDCK-IOWA cell lines as part of this project, to test the effects of the intron 1 elements: however, no fluorescent cells were detected (data not shown). Barbara Stevenson (another

member of the Edinburgh MMC group) created the vector pS1G for a different project. This vector fuses GFP to the C-terminus of CFTR: expression is driven by a P_{CMV} promoter. When this vector was transfected into the COS-7 cell line, fluorescent cells were detected, however the average fluorescence intensity of the transfected population was very weak (see Fig. 4.2).

Although GFP fluorescence is a property of the chromophore region of the protein, there is evidence that changes to other regions of the protein may affect the intensity and spectral profile of fluorescence. Specifically, there are reports that protein fusions may affect EGFP fluorescence (Cha et al, 2000). We believe this is what is happening with the pS1G and $_{PAC}$RC2EGFP vectors. When the pCFTRiresEGFP vector was transfected into the COS-7 cell line (Fig. 5.1), cells appeared to express EGFP. Although the average fluorescence intensity of the transfected population (X_f = 131.75) was reduced in respect to pEGFP-N (X_f = 4247.85), it was significantly above background levels (X_f = 4.01). Although it is not strictly appropriate to compare values between experiments, this value exceeded the fluorescence intensity of the pS1G vector (X_f ~97.14, as measured by the histogram method, Fig. 4.2). In light of this comparison, and the failure to detect EGFP expression from the $_{PAC}$RC2EGFP vector, the IRES approach was selected for the creation of additional vectors $_{PAC}$RC1iresEGFP and $_{PAC}$RC2iresEGFP.

These studies were unable to resolve the effects of the CFTR intron 1 DHS upon EGFP expression in gut cell lines. The transfection of control constructs pEGFP-N, $_{PAC}$RC2cmvEGFP, and pCFTRiresEGFP showed that these lines have a generally low transfection efficiency, that expression is reduced with large PAC vectors, and that translation from an internal ribosome entry site is inefficient in these cell lines. In creating additional reporter constructs, it will probably be necessary to use small plasmid vectors and/or find an alternative to the IRES system to create vectors with detectable levels of expression.

One potential way to overcome the low transfection efficiency with PAC vectors would be the production of immortalised cell lines. In fact, several stable cell lines were made by integrating the p1kbcfproEGFP construct into the COS-7 cell line, as part of this project. However, EGFP expression in these lines did not appear to be

copy-number dependent, and appeared to suffer from position effects (results not shown). YAC vectors containing more distal CFTR 5' regions, and 3' regions appear to exhibit copy-number dependent, position independent expression (Vassaux and Huxley, 1997); this may be a result of insulator/boundary elements surrounding the CFTR locus, or may be an intrinsic effect of large YAC sequences. Thus, it may be possible to analyse the effects of CFTR regulatory elements by creating permanent cell lines with the larger PAC vectors. Even if copy-number dependent, position independent expression were not achieved, these permanent cell lines could be useful in analysing the effects of pharmacological elements (such as forskolin and PMA) upon expression.

A fusion where GFP is linked to the N-terminus of CFTR has been created (Moyer et al, 1998): this fusion protein expresses strong fluorescence, and maintains CFTR channel activity. It would seem logical to duplicate this fusion in GCVs. This method is being pursued by other members of the Edinburgh MMC; however, it will be difficult to engineer vectors containing fusions to the CFTR N-terminus, using our current methods.

One paper by Ann Harris' group (Smith, AN et al, 1996) suggests that CFTR intron 1 does not require its endogenous location to affect expression, and that it is capable of upregulating expression from proximal 5' region of CFTR, where other CFTR genomic context elements are absent. In light of this information, a reasonable approach to assess the effects of CFTR intron 1 upon EGFP expression would be to insert this element into the p1kbcfproEGFP vector. This vector could be created by conventional cloning, by amplifying the intron 1 DHS region with the polymerase chain reaction, and then inserting this product into the remaining portion of the multiple cloning site (upstream of the proximal promoter), or into a restriction site downstream of the EGFP gene, in the p1kbcfproEGFP vector.

Two reports have shown that the intron 1 element can upregulate expression of the Luciferase reporter gene (Smith, AN et al, 1996; Mogayzel and Ashlock, 2000). EGFP reporter plasmids would give a more detailed picture of the dynamics of this upregulation, for the reasons described in Chapter 4.

It would be interesting to create EGFP constructs including other CFTR introns or 3' region, to test the effects of the various CFTR DHS elements upon expression. Finally, it would be useful to create additional constructs, with a cut-down version of the CFTR promoter. Other reports suggest that the DHS at –20.9 may upregulate CFTR expression by 60% in the Caco-2 cell line (Nuthall et al, 1999b). The methods used here were not sensitive enough to distinguish this relatively subtle increase in expression. It would be useful to create a cut-down vector, containing a small region encompassing the –20.9 DHS, linked to the proximal promoter in a small construct, to further evaluate the effects of this DHS upon expression.

EGFP has many other applications in the study of CFTR function, which are more divorced from this project. One example relates to the question of whether the CFTR functional channel is a dimer. Fluorescence resonance energy transfer (FRET) is a powerful technique to study protein interactions (Pollock and Heim, 1999). In this method, fusion proteins are made between the proteins in question and differently coloured GFP variants (with different excitation and emissions spectra –see Section 1.3.1). The proteins are then co-transfected into cell lines. If the proteins in question are not interacting, than each fluorescent protein will only glow when excited with light in the appropriate region of the spectrum corresponding to its *own* excitation spectra. However, if the proteins in question are interacting, then the close proximity will result in "cross-talk" between the fused GFPs, and both GFP variants will fluoresce when only one is excited. Thus, CFTR could be fused with differently coloured GFP molecules, and this FRET technique could be used to determine whether they are interacting, e.g. forming dimers.

GFP-CFTR fusions will also be valuable in studying CFTR trafficking within the cell: this approach is being pursued by other groups (Moyer et al, 1998; Moyer et al, 1999; Loffing-Cueni et al, 2001).

7.2.3 Size effects/Plasmid mixing effects

Additional transfection studies could be used to investigate the vector size effect (e.g. differential expression in PACs and plasmids), and the effects of plasmid mixing. It would be useful to test a large BAC vector containing the EGFP reporter gene under

the control of a P_{CMV} promoter, to see if it is large vector size *per se* which reduces expression. Conversely, small plasmid vectors could be created containing PAC backbone elements to rule out the possibility that it is these which are reducing expression.

The plasmid mixing experiments unveiled an unexpected interaction: that diluting a reporter plasmid with an anonymous plasmid appears to increase its transfection efficiency. This effect cannot yet be explained, and it would be useful to perform further experiments to investigate the mechanics of this process. Firstly, it will be necessary to repeat this experiment with different reporter genes, and different transfection regents (so far only LID and SID have been tried), to determine whether this phenomenon is universal. It would also be interesting to perform plasmid mixing with fluorescently labelled DNA vector (such as pGeneGrip) to determine the fate of the DNA after it has entered the cell (by microscopy). It would be interesting to investigate the effects of mixing the reporter plasmid with different forms of DNA, such as mammalian chromosomal DNA (this might mask the presence of the bacterial plasmid and thereby increase transfection efficiency). This effect certainly appears promising, and warrants further investigation.

7.2.4 Primary Cultures

The sheep primary tracheal air-interface cultures (described in Chapter 6) demonstrate features of the endogenous trachea and are a useful *ex vivo* model. However, further experiments are required to characterise these cultures. For example, antibody-staining experiments should be performed to fully characterise the cell types present in these cultures.

One attempt was made to identify the cell types which are being transfected by mechanical FACS sorting (e.g. separating the population of cells which are expressing EGFP). The cells were subsequently applied to a poly-L-lysine coated slide, fixed, and stained with hemotoxylin/eosin to visualise cell structures. This attempt failed, as the cells were lysed at some point during the process (only cell debris remained). It is likely that this was a result of using a FACSCalibur to sort the cells: this cytometer sorts cells mechanically and does not preserve cell viability. The

more sophisticated FACSVantage cytometer, which sorts the cells magnetically and maintains their viability, might produce better results.

In retrospect, it might have been possible to determine the identity of the expressing cells more simply with fluorescence microscopy. The pEGFP-N plasmid produced reasonable transfection efficiency, and the average fluorescence intensity of the transfected population was significantly above background, such that these cells could probably be distinguished by microscopy. However, it might be difficult to determine whether it is the underlying, or top layer of cells (in the pseudostratified bilayer) expressing EGFP with this method. It would probably be necessary to combine this with cell staining, or immunohistochemistry (e.g. for markers of cell identity, or antibodies that recognise structures such as cilia), to determine the identity of the transfected cells.

In the lung, CFTR expression is not uniform amongst epithelial cells, but appears to be higher in certain subpopulations of cells, in both the submucosal glands and the airways (Engelhardt, 1992 and 1994). Some of these cell types may be represented in the primary air-interface cultures. Reporter constructs containing the EGFP gene would be ideally suited to test whether CFTR promoter elements can restrict expression to specific cell types.

Ann Doherty and Peter Thorpe, members of the Edinburgh MMC group, are pursuing some of these experiments relating to primary cultures. In addition, Ann Doherty is attempting to establish similar cultures from lower portions of the airway, such as the bronchus, which may prove to be the target for gene therapy. Ultimately, other *ex vivo* primary culture methods and *in vivo* animal experiments will be required to refine gene therapy vectors prior to commencing a clinical trial.

7.3 Closing remarks

This project investigated the use of EGFP as a reporter to evaluate CFTR regulatory elements, in gene therapy vectors. We can make some general conclusions about the strengths and weaknesses of the methods used in this study.

Vector size dramatically affected transfection efficiency in cell lines: this made it difficult to assess the effects of regulatory elements upon expression by comparing expression from small plasmid and large PAC vectors. This study highlighted the limitations of working with transformed cell lines, such as COS-7, when the end purpose is to evaluate vectors designed for use *in vivo*. The air-interface *ex vivo* system described here is a more accurate approximation of the native epithelium. Studies in air-interface cultures revealed the unexpected observation that the effect of vector size was less pronounced than in e.g. COS-7 cells. There is however a clear and substantial effect of the promoter (heterologous vs. autologous). An inherent problem with all such comparative studies is the day-to-day variation in transfection efficiency. Consequently, it was necessary to transfect vectors simultaneously to allow a comparison of their effects. This obligation restricted the size and scale of experiments.

Nevertheless, this report has shown that EGFP can be used as a reporter of weak promoters, such as the CFTR proximal 5' region. The dual dotplot method has proved to be a superior method of analysing FACS data, creating a very sensitive and quantitative assay for EGFP expression. Finally, an interesting interaction was uncovered: that dilution of a reporter gene does not necessarily reduce its transfection efficiency.

EGFP has proved to be an excellent reporter of CFTR promoter activity in gene therapy vectors. Further studies utilising this reporter gene are warranted.

Bibliography

Abdallah, B et al (1996) A powerful nonviral vector for in vivo gene transfer into the adult mammalian brain: Polyethylenimine. *Hum. Gene Ther.* **7:** 1947 - 1954

Alton, EW et al (1993) Non-invasive liposome-mediated gene delivery can correct the ion transport defect in cystic fibrosis mice. *Nature Genetics* **5:** 135 – 142

Alton EW et al (1999) Cationic lipid-mediated CFTR gene transfer to the lungs and nose of patients with cystic fibrosis: a double-blind placebo-controlled trial. *Lancet* **353:** 947 - 954

Albano, CR (1998) Green Fluorescent Protein as a real time quantitative reporter of heterologous protein production. *Biotechnol. Prog.* **14:** 351 – 354

Anderson, DH (1938) Cystic fibrosis of the pancreas and its relation to celiac disease. *Am J. Dis. Child* **56:** 344 - 402

Anderson and Hodges (1946)

Anderson, MT et al (1996) Simultaneous fluorescence-activated cell sorter analysis of two distinct transcriptional elements within a single cell using engineered green fluorescent proteins *PNAS* **93:** 8505 – 8511

ATCC website http://www://atcc.org

Baker, A et al (1997) Polyethylenimine (PEI) is a simple, inexpensive an effective reagent for condensing and linking DNA to adenovirus for gene delivery. *Gene Therapy* **4:** 773 - 782

Bargon, J et al (1998) Modulation of cystic fibrosis transmembrane conductance regulator gene-expression by elevation of intracellular cyclic AMP *Eur. J. Med. Res.* **3:** 256 – 262

Bartlett, JS and Samulski, RJ (1998) Fluorescent viral vectors: A new tool for the pharmacological analysis of gene therapy. *Nature Medicine* **4:** 635 – 637

Baudouin-Legros, K et al (2000) Modulation of CFTR gene expression in HT-29 cells by extracellular hyperosmolarity. *Am. J. Physiol. Cell Physiol.* **278:** C49 – 56

Beardsley, T (1990) Clearing the airways: Cystic Fibrosis may be treated with gene therapy *Sci. Am.* **263:** 28,30,33

Bienvenu, T et al (1995) Three novel sequence variations in the 5' upstream region of the cystic fibrosis (CFTR gene): two polymorphisms and one putative molecular defect *Hum. Genet.* **95:** 698 – 702

Blaauw, M et al (2000) Efficient control of gene expression by a tetracycline-dependent transactivator in single Dictyostelium discoideum cells. *Gene* **252:** 71 - 82

Boucher, RC (1994) Human airway ion transport. Part I. *Am. J. respir. Crit Care Med.* **150:** 271 – 281

Boussif, O et al (1995) A versatile vector for gene and oligonucleotide delivery into cells in culture and *in vivo*: Polyethylenimine. *PNAS* **92:** 7297 - 7301

Boyd, AC et al (1999a) Insertion of natural intron 6a-6b into a human cDNA-derived gene therapy vector for cystic fibrosis improves plasmid stability and permits facile RNA/DNA discrimination. *J. Gene Med.* **1:** 312 - 321

Boyd, AC et al (1999b) Construction and characterisation of genomic context vectors for CF gene therapy. (abstract) *Ped. Pulm.* **S19:** 237.

Boyd and Porteous (1997) PCR-Generated Cross-over Linkers for Site-Directed Mutagenesis. *Biotechniques* **23:** 827 - 830

Bragzoni, A et al (1999) Comparison between cationic polymers and lipids in mediating systemic gene delivery to the lungs. *Gene Therapy* **6:** 1995 – 2004

Braun, R et al (1998) Compatibility of plasmids expressing different antigens in a single DNA vaccine formulation. *J. Gen. Virol.* **79:** 2965 - 2970

Bruer, W et al (1992) Induction of expression of the cystic fibrosis transmembrane conductance regulator *J. Biol. Chem.* **267:** 10465 - 10469

Bubien, J et al (1990) Cell cycle dependence of chloride permeability in normal and cystic fibroisis lymphocytes. *Science* **248:** 1416 – 1419

Cancer Research UK FACS laboratory website: http://www.icnet.uk/axp/facs/davies/flow.html

Cantin, A (1995)*Cystic fibrosis lung inflammation: early, sustained and severe.* Am J Respir Crit. Care Med. **151:** 939 – 941

Caplen, NJ et al (1995) Liposome mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis. *Nature Medicine* **1:** 39 - 46

Carey, M and Smale, ST *Transcriptional Regulation in Eukaryotes: concepts, strategies and Techniques* © Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY 2000

CF mutations database: http://www.genet.sickkids.on.ca/cftr/

CF Foundation website: http://www.cff.org/

Cha, HJ et al (2000) Observations of green fluorescent protein as a fusion partner in genetically engineered Escherichia coli: monitoring protein expression and solubility. *Biotechnol. Bioeng.* **67:**565-74

Chalfie, M et al (1994) Green fluorescent protein as a marker for gene expression. *Science* **263:** 802 – 805

Chen, JH et al (1989) A cAMP-regulted chloride channel in lymphocytes that is affected in Cystic Fibrosis. *Science* **243:** 657 – 660

Chiocchetti, A et al (1997) Green fluorescent protein as a reporter of gene expression in transgenic mice. *Biochem. Biophys. Actica* **1352:**193-202

Chou, JL et al (1991) Characterization of the promoter region of the cystic fibrosis transmembrane conductance regulator gene. *J. Biol. Chem.* **266:** 24471 - 24476

Chou, Q et al (1992) Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* **20:** 1717 – 1723

Chung, JH et al (1997) Characterisation of the chicken beta-globin insulator. *PNAS* **94:** 575 – 580

Clarke, LL et al (1002) Defective epithelial chloride transport in a gene-targeted mouse model of cystic fibrosis. *Science* **257:** 1125 – 1128

Clontech website http://www.clontech.com

Cohen, CJ et al (2001) The cocksackievirus receptor in a transmembrane component of the tight junction. *PNAS* **98:** 15191 – 15196

Colledge, WH et al (1995) Generation and characterization of a delta F508 cystic fibrosis mouse model. *Nat. Genet.* **10:** 445 - 452

Cormack BP et al (1997) Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans. *Microbiology* **143:** 303 – 311

Coyne, CB et al (2000) Enhanced epithelial gene transfer by modulation of tight junctions with sodium caprate. *Am. J. Respir. Cell. Mol. Biol.* **23:** 602-609

Crawford, I et al (1991) Immunocytochemical localization of the cystic fibrosis gene product CFTR *PNAS* **88:** 9262 – 9266

Crystal, RG et al (1994) Administration of an adenovirus containing human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis. *Nature Genetics* **8:** 42-50 (1994)

Davidson, DJ et al (1995) Lung disease in the cystic firosis mouse exposed to bacterial pathogens. *Nature Genetics* **9:** 351 - 357

Davidson, DJ et al (2000) A primary culture model of differentiated murine tracheal epithelium. *Am. J. Physiol Lung Cell Mol Physiol.* **297:** L766 – L778

Davidson, DJ and Dorin, JR (2001) The CF mouse:an inportant tool for studying cystic fibrosis. Exp. Rev. Mol. Med.
http://www-ermm.cbcu.cam.ac.uk/01002551h.htm

Davidson, DJ and Porteous, DJ (1998) The genetics of cystic firosis lung diesease. *Thorax* **53:** 389 - 397

Davidson, H et al (2000) Genomic sequence analysis of fugu rubripes CFTR and flanking genes in a 60 kb region conserving sytney with 800 kilobases of human chromosome 7. *Genome Research* **10:**1194 –1203

Davies, JC et al (1997) CFTR gene transfer reduces the binding of pseudomonas aeruginosa to Cystic Fibrosis respiratory epithelilum. *Am. J. Resp. Cell Mol. Biol.* **16:** 657 – 663

Davies, JC et al (1998) Prospects for gene therapy for cystic fibrosis. *Molecular Medicine Today* **4:** 292 – 299

Delaney, DJ et al (1996) Cystic fibrosis mice carrying the missense mutation G551D replicate human genotype-phenotype correlations. *EMBO J.* **15:** 955 - 963

Dharmsathaphorn, K et al (1984) A human colonic tumor cell line that maintains vectorial electrolyte transport. *Am. J. Physiol.* **246:** G204-G208

Dickinson, P et al. (1998) Generation of a CF mutant mouse possessing the G480C mutation. *In 22nd European CF Conference Berlin Book of Abstracts*, PS7.14, p. 143

Doorninck, J et al (1995) A mouse model for the cystic fibrosis deltaF508 mutation. *EMBO* **14:** 4403 - 4411

Dorin, JR et al (1992) Cystic fibrosis in the mouse by targeted insertional mutagenesis. *Nature* **357:** 211 – 215

Dorin, JR et al (1996) A demonstration using mouse models that successful gene therapy for cystic fibrosis requires only partial gene correction. *Gene Therapy* **3:** 797- 801

Douglas, J and Curiel, D (1998) Gene therapy for inherited inflammatory and infectious disease of the lung. *Medscape Respir. Care* **2:**
http://www.medscape.com/Medscape/Respir.../v02.n03/mrc4574.curi/mrc4574.curi.html

Drumm, ML et al (1990) Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer. *Cell* **62:** 1227-1233
Dunn, AK and Handelsman, J (1999) A vector for promoter trapping in Bacillus cereus. *Gene* **226:** 297 – 305

Ellsworth, RE et al (2000) Comparative genomic sequence analysis of the human and mouse cystic fibrosis transmembrane conductance regulator genes. *PNAS* **97:** 1172 – 1177

Engelhardt, JF et al (1992) Submucosal glands are the predominant site of CFTR expression in the human bronchus. *Nature Genetics* **2:** 240 – 247

Engelhardt, JF et al (1994) Expression of the cystic fibrosis gene in adult human lung. *J. Clin. Invest.* **93:** 737 – 739

Ensembl database http://www.ensembl.org/

Flotte, TR et al (2001) Gene therapy in Cystic Fibrosis. *Chest* **120:** 124S –131S

Fogh, J et al (1977a) Absence of HeLa cell contamination in 169 cell lines derived from human tumors. *J. Natl. Cancer Inst.* **58:** 209-214

Fogh, J et al (1977b) One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J. Natl. Cancer Inst.* **59:** 221-226

Frizzel, RT (1995) Functions of the cystic fibrosis transmembrane conductance regulator gene. *Am. J. Respir. Crit. Care Med.* **151**: S54 – S58

Gaiso, M (1998) Cystic fibrosis: A role for gene therapy? *Medscape online article* http://www.medscape.com/Medscape/featur...98/05.98/med0527.gais/med0527.gais.html

Gaush, CR et al (1966) Characterization of an established line of canine kidney cells (MDCK). *Proc. Soc. Exp. Biol. Med.* **122:** 931-935

Gill, DR et al (1977) A placebo controlled study of liposome-mediated gene transfer to the nasal epithelium of patients with cystic fibrosis. *Gene Therapy* **4:** 199 – 209

Goddard, CA et al (1997) A second dose of a CFTR cDNA-liposome complex is as effective as the first dose in restoring cAMP-dependent chloride secretion to null CF mice trachea. *Gene Therapy* **4:** 1231-1236

Grade, K et al (1994) Identification of three novel mutations in the CFTR gene using temperature-optimized non-radioactive conditions for SSCP analysis. *Hum. Genet.* **94:** 154 - 158

Grosveld, F et al (1997) Position independent, high level expression of the human beta-globin gene in transgenic mice. *Cell* **51:** 975 – 985

Gruenert, DC et al (1995) Culture and transformation of human airway epithelial cells. *Am. J. Physiol.* **268:** L347 – 3460

Guggino, WB (1999) Cystic Fibrosis and the salt controversy. *Cell* **96:** 607 – 610

Harris, A (1997) Towards an Ovine model of cystic fibrosis. *Human Molec. Genet.* **6:** 2192 – 2194

Hart, SL et al (1995) Gene delivery and expression mediated by an integrin binding peptide. *Gene Therapy* **2:** 552 – 554

Hart, SL et al (1998) Lipid mediated enhancement of transfection by a novel integrin-targeting vector. *Human Gene Therapy* **9:** 575 - 585

Hart, SL (1999) Integrin-mediated vectors for gene transfer and therapy. *Curr. Opin. Mol. Ther.* **1:** 197 – 203

Hasty, P et al. (1991) Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells [published erratum appears in Nature 1991 Sep 5; 353(6339): 94]. *Nature* **350:** 243-246

Heeckeren, A et al (1997) Excessive inflammatory response of cystic fibrosis mice to bronchopulmonary infection with Pseudomonas aeruginosa. *J. Clin. Invest.* **100:** 2810 - 2815

Heid, CA et al (1996) Real time quantitative PCR. *Genome Res.* **6:** 986 – 994

Higgins, CF et al (1990) Binding protein-dependent transport systems. *J. Bioenerg. Biomembr.* **22:** 571 – 592

Ho, PJ and Thein, SL (2000) Gene Regulation and deregulation: a beta globin perspective *Blood Rev.* **14:** 78 - 93

Hollingsworth, MA et al (1994) A nuclear factor that binds purine-rich, single-stranded oligonucleotides derived from s1-sensitive elements upstream of the CFTR gene and the MUC1 gene. *Nucl. Acid. Res.* **22:** 1138 - 1146

Holmes, AR et al (1999) Intracellular compartmentalisation of DNA fragments in cultured airway epithelial cells mediated by cationic liposomes. *Pharm. Res.* **16:** 1020 – 1025

Horowitz, B et al (1993) Alternative splicing of CFTR Cl- channels in heart. *Am. J. Physiol.* **264:** H2214 – 2220

Hyde, SC et al (1993) Correction of the ion transport defect in cystic fibrosis transgenic mice by gene therapy. *Nature* **362:** 250 – 255

Imler, JL et al (1996) Targeting cell-specific gene expression with adenovirus containing the LacZ gene under the control of the CFTR promoter. *Gene Therapy* **3:** 49 – 58

Ismailov, II et al (1996) Regulation of epithelial sodium channels by the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **271:** 4725 – 4732

Jackson, RJ et al (1990) The novel mechanism of initiation of picornvirus RNA translation. *TIBS* **15:** 477 - 483

Jenkins, RG et al (2000) An integrin-targeted non-viral vector for pulmonary gene therapy. *Gene Therapy* **7:** 393 - 400

Johnson, LG et al (1992) Efficiency of gene transfer for restoration of normal airway epithelial function in cystic firosis. *Nature Genetics* **2:** 21 - 25

Joris, L and Quinton, PM (1989) Evidence for electrogenic Na-glucose cotransport in tracheal epithelium. *Pflugers Arch.* **415:** 118-120

Kain, SR et al (1995) Green fluorescent protein as a reporter of gene expression and protein localization. *BioTechniques* **19:** 650 – 655

Kälin, N et al (1999) DeltaF508 CFTR protein expression in tissues from patients with Cystic Fibrosis. *J. Clin. Invest.* **103:** 1379 – 1389

Kartner, N et al (1992) Mislocalization of Delta F508 CFTR in cystic fibrosis sweat gland. *Nature Genetics* **1:** 321 – 327

Kelley, TJ and Drumm, ML (1998) Inducible nitric oxide synthase expression is reduced in cystic fibrosis murine and human airway epithelial cells. *J. Clin. Invest.* **102:**1200 – 1207

Kerem, B et al (1989) Identification of the Cystic Fibrosis gene: genetic analysis. *Science* **245:** 1073 – 1080

Knowles, MR et al (1995) Pharmacological modulation of salt and water in the airway epithelium in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **151:** S65 – S69

Koh, J et al (1993) Characterization of the Cystic Fibrosis Transmembrane Conductance Regulator promoter region. *J. Biol. Chem.* **268:** 15912 - 15921

Kolossov, E et al (1998) Functional characteristics of ES-cell derived cardiac precursor cells identified by tissue-specific expression of green fluorescent protein. *J. Cell Biol.* **143:** 2045 – 2056

Kondo, M et al (1997) Increased oxidative metabolism in cow tracheal epithelial cells cultured at air interface. *Am. J. Respir Cell Mol. Biol.* **16:** 62 – 68

Korst, RJ et al (1995) Gene Therapy for the respiratory manifestations of cystic firosis *Am. J. Respir. Crit. Care Med.* **151:** S75 - S87

Kulka, M et al (2002) Expression and functional characterization of CFTR in mast cells. *J. Leuk. Biol.* **71:** 54 – 62

Larson, JE et al (1997) Reversal of cystic fibrosis phenotype in mice by gene therapy in utero *Lancet* **349:** 619 – 620

Levenson, W et al (1998) Internal riosome entry site-containing retroviral vectors with green fluorescent protein and drug rsistance markers. *Human Gene Therapy* **9:** 1233 – 1236

Levesque, PC et al (1992) Expression of cystic fibrosis transmembrane conductance regulator Cl- channels in the heart. *Circ. Res.* **71:** 1002 – 1007

Li, S et al (1999) Transcriptional repression of the cystic fibrosis transmembrane conductance regulator gene, mediated by CCAAT displacement protein/cut homolog, is associated with histone deacetylation. *J Biol. Chem.* 274: 7803 – 7815

Li, CD et al (2001) A monomer is the minimal functional unit required for channel and ATPase activity of the cystic fibrosis transmembrane conductance regulator. (abstract) *Ped. Pulm. Supplement* **22:** 176

Lin, PY and Gruenstein, E (1987) Identification of a defective cAMP-stimulated Cl-channel in Cystic Fibrosis Fibroblasts. *J. Bio. Chem.* **262:** 15345 – 15347

Lo, W et al (1998) Making genes green: Creating greenf fluorescent protein (GFP) fusions with blunt-end PCR products. *Biotechniques* **25:** 94 – 98

Loffing-Cueni, D et al (2001) Trafficking of GFP-tagged DeltaF508-**CFTR** to the plasma membrane in a polarized epithelial cell line. *Am. J. Cell Physiol.* **281:** C1889 – 1897

Lukacs, GL et al (1993) The delta F508 mutation decreases the stability of cystic fibrosis transmembrane conductance regulator in the plasma membrane. *J. Biol. Chem.* **29:** 21592 – 21598

Manson, AL et al (1997) Complementation of null CF mice with a human CFTR YAC transgene *EMBO J.* **16:** 4238 – 4249

Marshall, J et al (1995) The jellyfish green fluorescent protein: A new tool for studying ion channel expression and function. *Neuron* **14:** 211 – 215

Mateu, E et al (2002) Can a place of origin of the main CF mutations be identified? *Am J. Human Gent.* **70:** 257 - 264

Matthews, RP and McKnight PS (1996) Characterization of the cAMP response element of the cystic fibrosis transmembrane conductance regulator gene promoter. *J. Biol. Chem.* **271:** 31869 - 31877

McBride, S et al (2000) Gene transfer to adult human lung tissue ex vivo. *Gene Therapy* **7:** 675 – 678

McCray, PB et al (1992) Localisation of Cystic Fibrosis Transmembrane conductance regulator in human fetal lung tissue by in situ hybridisation. *J. Clin. Invest.* **90:** 619 - 625

McDonald, RA et al (1995)Basal Expression of the cystic fibrosis transmembrane conductance regulatorgene is dependent on protein kinase A activity. *PNAS* **92:** 7560 - 7564

McLachlan, G et al (1995) Evaluation in vitro and in vivo of cationic liposome-expression construct complexes for cystic fibrosis gene therapy. *Gene Therapy* **2:** 614 – 22

McLachlan, G et al (1996) Laboratory and clinical studies in support of cystic fibrosis gene therapy using pCMV-CFTR-DOTAP. *Gene Therapy* **3:** 1113 – 1123

Mickle, JE et al (1998) A mutation in the cystic fibrosis transmembrane conductance regulator gene associated with elevated sweat chloride concentrations in the absence of cystic fibrosis. *Human Mol.* Genet. **7:** 729 – 735

Middleton, PG and Alton, EW (1998) Gene therapy for cystic fibrosis: which postman, which box? *Thorax* **53:** 197 - 199

Mogayzel, PJ and Ashlock, MA (2000) CFTR intron 1 increases luciferase expression driven by CFTR 5' flanking DNA in a yeast artificial chromosome. *Genomics* **64:** 211 – 215

Mohammed-Panah, R et al (1998) Hyperexpression of recombinant CFTRin heterologous cells alters its physiological properties. *Am. J. Physiol.* **274:** C310 – 318

Moyer, BD et al (1998) Membrane trafficking of the cystic fibrosis gene product, cystic fibrosis transmembrane conductance regulator, tagged with green fluorescent protein in Mandin-Darby canine kidney cells. *J. Biol. Chem.* **273:** 21759 – 21768

Moyer, BD et al (1999) Butyrate increases apical membrane **CFTR** but reduces chloride secretion in MDCK cells. *Am. J. Physiol.* **277**: F271 – 276

Nuthall, HN et al (1999a) Analysis of DNase I hypersensitive sites at the 3' end of the cystic fibrosis conductance regulator gene (CFTR). *Biochem. J.* **341:** 601 – 611

Nuthall, HN et al (1999b) Ananlysis of a DNase I hypersensitive site located –20.9 kb upstream of the CFTR gene. *Eur. J. Biochem* **266:** 431 - 433

O'Neal, WK et al (1993) A severe phenotype in mice with a duplication of exon 3 in the cystic fibrosis locus. *Hum. Mol. Genet.* **2:** 1561-1569

Ornoy, A et al (1987) Pathological confirmation of cystic fibrosis in the fetus, following prenatal diagnosis. *Am. J. Med. Genet.* **28:** 935 – 947

Pasyk, EA and Foskett, JK (1997) Cystic fibrosis transmembrane conductance regulator-associated ATP and adenosine 3'-phosphate 5'-phosphosulfate channels in endoplasmic reticulum and plasma membranes. *J. Biol. Chem.* **272:** 7746 - 7751

Persons, DA et al (1998) Use of the green fluorescent protein as a marker to identify and track genetically modified hematopoietic cells. *Nature Medecine* **4:** 1201 – 1205
Phillipson, G (1998) Cystic Fibrosis and reproduction. *Reprod. Fertil. Dev.* **10:**113 – 119

Phylactides, M et al (2002) Evaluation of potential regulatory elements identified as DNase I hypersensitive sites in the CFTR gene. *Eur. J. Biochem.* **269:** 553 - 559

Pier, GB et al (1997) Cystic fibrosis transmembrance conductance regulator is an epithelial cell receptor for clearnce of Pseudomonas aeruginosa from the lung. *PNAS* **94:** 12088 – 12093

Pier, GB (2000) Role of the cystic fibrosis transmembrane conductance regulator in innate immunity to Pseudomonas aeruginosa infections *PNAS* **97:** 8822 – 8828

Pilewski, JM and Frizzel, RA (1999) Role of CFTR in airway disease. *Physiol. Rev.* **79:** S215 - 255

Pines, J (1995) GFP in mammalian cells. *Trends. Genet.* **11:** 326 - 327

Pittman, N et al (1995) Transcription of cystci fibrosis conductance regulator requires a CCAT-like element for both basal and cAMP-mediated rehulation *J. Biol. Chem.* **270:** 28848 – 28857

Pollock, BA and Heim, R (1999) Using GFP in FRET-based applocations. *Trends Cell Biol.* **9:** 57 – 60

Porteous, DJ et al (1997) Evidence for safety and efficacy of DOTAP cationic liposome mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis. *Gene Therapy* **4:** 210 – 218

Porteous, DJ and Davidson, D (1997) Cystic fibrosis lung infection cleared up? *Nature Medicine* **3:** 1317-1318

Quinton, PM (1983) Chloride impermability in cystic fibrosis. *Nature* **301:** 421 – 422

Raghuram, V et al (2001) Functional CFTR channel is a homo-dimer. Abstract in *Ped. Pulm.* **S22:** 176

Ratcliff, R et al (1993) Production of a severe cystic fibrosis mutation in mice by gene targeting. *Nat. Genet.* **4:** 35-41

Rich, DP et al (1990) Expression of cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. *Nature* **347:** 358-363

Rich, DP et al (1997) Effect of deleting the R domain on CFTR-generated chloride channels. *Science* **253:** 205 – 207

Richman-Eisenstat, J (1996) Cytokine Soup: making sense of inflammation in cystic fibrosis. *Ped. Pulm.* **21:** 3-5

Riordan, JR et al (1989) Identification of the Cystic Fibrosis gene: cloning and characterisation of complementary DNA. *Science* **245:** 1066 – 1073

Roche website: http://www.roche-applied-science.com/

Rommens, JM et al (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245:** 1059 – 1065

Rosenfeld, MA et al (1992) In vivo transfer of the human cystic fibrosis conductance regulator gene to the airway epithelium. *Cell* **68:** 143 – 155

Rowntree, RK et al (2001) An element in intron 1 of the CFTR gene augments intestinal expression in vivo. *Hum. Molec Genet.* **10:** 1455 – 1464

Rozmahel, R et al (1996) Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. *Nature Genetics* **12:** 280 – 287

Saffery, R and Choo, KH (2002) Strategies for engineering human chromosomes with therapeutic potential *J. Gene Medicine* **4:** 5 - 13

Saint B.V. website http://www.Saint-Mix.com

Sauer, B Methods of Enzymology; 1993, Vol. 225, 890-900

Schiavi, SC et al (1996) Biosynthetic and growth abnormalities are associated with high-level expression of CFTR in heterologous cells. *Am. J. Physiol.* **270:** C341 – 351

Schnieke, AE at al (1997) Human factor IX transgenic shepp produced by transfer of nuclei from transfected fetal fibroblasts. *Science* **278:** 2130 – 2133

Scott, ES et al (2000) A murine tracheal culture system to investigaet parameters affecting gene therapy for cystic fibrosis. *Gene Therapy* **7:** 612 – 618

Smith, AN et al (1995) Characterisation of DNase I hypersensitive sites in the 120 kb 5' to the CFTR gene. *Biochem. Biophys. Res. Comm.* **271:** 274 – 281

Smith, AN et al (1996) A regulatory element in Intron 1 of the cystic fibrosis transmembrane conductance regulator gene. *J. Biol. Chem.* **271:** 9947 – 9954

Smith, DJ et al (2000) Multiple potential intragenic regulatory elements in the CFTR gene. *Genomics* **64:** 90 – 96

Smith, JJ et al (1996) Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* **85:** 229 – 236

Snouwaert, JN et al (1995) A murine Model of Cystic Fibrosis. *Am.J. Respir. Crit. Care Med.* **151:** S59 - S64

Stanford 5-minute guide to flow cytometry website:
http://facs.stanford.edu/5MinuteGuide/5MinuteGuide.htm

Stauber, RH et al (1998) Development and applications of enhanced green fluorescent protein mutants. *Biotechniques* **24:** 462 – 471

Stripecke, R et al (1999) Immune response to green fluorescent protein: implications for gene therapy. *Gene Therapy* **6:** 1305 – 1312

Stutts, MJ et al (1993) Functional consequences of heterologous expression of the cystic fibrosis transmembrane conductance regulator in fibroblasts. *J. Biol. Chem.* **268:** 20653 – 20658

SurfStat Statistical tables website:
http://math.uc.edu/~brycw/classes/148/tables.htm#normal

Subramanian, S and Srienc, F (1996) Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. *J. Biotechnol.* **49:** 137 – 151

Tebbutt, SJ et al (1995) Molecular analysis of the ovine cystic fibrosis transmembrane conductance regulator gene. *PNAS* **92:** 2293 – 2297

Thomaisis, TS and Arey, JB (1963) The intestinal lesion in cystic fibrosis of the pancreas. *J. Pediatr.* **63:** 444 – 453

Thorpe, PH et al (2002) Functional correction of episomal mutations with short DNA fragments and RNA-DNA oligonuclotides. *J. Gene Medicine* **4:** 195 - 204

Tousson, A et al (1998) Characterisation of CFTR expression and chloride channel activity in human endothelia. *Am. J. Physiol.* **275:** C1555 – C1564

Traber, DL and Traber, LD (1989) Sheep as a cardiopulmonary model. *Prog. Clin. Biol. Res.* **299:** 253 – 263

Trapnell, BC et al (1991a) Down-regulation of Cystic Fibrosis Gene mRNA Transcript levels and Induction of the Cystic Fibrosis Chloride Secretory Phenotype in Epithelial cells by Phorbol Ester. *J. Biol. Chem.* **266:** 10319 - 10323

Trapnell, BC et al (1991b) Expression of the cystic fibrosis transmembrane conductance regulator gene in the repsiratory tract of normal individuals and individuals with cystic fibrosis. *PNAS* **88:** 6565 – 6569

Trezise, AE and Buchwald, M (1991) In vivo cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature* **353:** 434 – 437

Tsui, LC et al (1985) Cystic fibrosis locus defined by a genetically linked polymorphic marker DNA. *Science* **230:** 1054 – 1057

Tsui, LC (1995) The cystic fibrosis transmembrane conductance regulator gene. *Am. J. Crit. Care Med.* **151:** S47 – S53

Valman, HB and France, NE (1969) The vas deferens in cystic fibrosis. *Lancet* **II:** 566 – 567

Vassaux, G et al (1997) Copy number-dependent expression of a YAC-cloned human CFTR gene in a human epithelial cell line. *Gene Ther*apy **4:** 618 – 623

Vassaux, G and Huxley, C (1997) A dicistronic construct allows easy detection of human CFTR expression from YAC DNA in human cells. *Nucleic Acids Res.* **25:** 4167 – 4168

Verlingue, C (1998) Absence of mutations in the interspecies conserved regions of the CFTR promoter region incystic fibrosis (CF) and CF related patients. *J. Med. Genet.* **35:** 137 – 140

Viallet, J et al (1994) Characterization of human bronchial epithelial cells immortalized by the E6 and E7 genes of human papillomavirus type 16. *Exp. Cell Res.* **212:** 36-41

Walmsley, M et al (1998) Primer extension analysis of mRNA. *Methods Mol. Biol.* **86:** 187 - 193

Wei, LY et al (1995) Overexpression of the cystic fibrosis transmembrane regulator in NIH 3T3 cells lowers the membrane potential and inracellular pH and confers a multidrug resistance phenotype. *Biophys J.* **69:** 883 – 895

Wilmut, I et al (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature* **385:** 810 – 813

Wilson, C et al (1990) Position effects on eukaryotic gene expression. *Annu. Rev. Cell Biol.* **6:** 679 – 714

Winter, MC and Welsh, MJ (1997) Stimulation of CFTR activity by its phosphorylated R domain. *Nature* **389:** 294 – 296

Wong, KR et al (1999) Molecular and functional distributions of chloride conductance in rabbit ventricle. *Am. J. Physiol.* **277:** H1403 – 1409.

Wood, WG (1996) The complexities of beta globin gene regulation. *Trends Genet.* **12:** 204 – 206

Yang, F et al (1996) The molecular structure of green fluorescent protein. *Nature Biotechnology* **14:** 1246 – 1251

Ye, L et al (2001) Creation of a sudden death mouse model by abberrant CFTR expression. Abstract in *Ped. Pulm.* **S22:** 226

Yoshimura, K et al (1991a) Expression of the cystic fibrosis transmembrane conductance regulator gene in cells of non-epithelial origin. Nucl.Acids. Res. **19:** 5417 - 5423, (1991a)

Yoshimura, K et al (1991b) The Cystic Fibrosis gene has a" housekeeping" –type promoter and is expressed at low levels in cells of epithelial origin. *J. Biol. Chem.* **266:** 9140 – 9144

Zabner, J et al (1993) Adenovirus-mediated Gene transfer transiently corrects the chloride defect in nasal epithelia of pateints with cystic fibrosis. *Cell* **75:** 207-216

Zeiher, BG et al (1995) A mouse model for the DeltaF508 allele of cystic fibrosis *J. Clin. Invest.* **96:** 2051 - 2064

Zerhusen, BD et al (2001) Two tightly coupled conductance pores for chloride ions formed by a dimeric structure of CFTR. Abstract in *Ped. Pulm.* **S22:** 177

Zhang, L et al (2001) Respiratory syncytial virus infects ciliated cells of airway epithelium via the luminal membrane. Abstract in *Ped. Pulm.* **S22**: 244

Zhang, Y et al (1998a) A new logic for DNA engineering using recombination in E. coli. *Nat. Genet.* **20:** 123 – 128

Zhang, Y et al (1998b) Vector-specific complementation profiles of two independent primary defects in Cystic Fibrosis airways. *Human Gene Therapy* **9:** 635 - 648

Zuelzer, WW and Newton, WA (1969) The pathogenesis of fibrocystic disease of the pancreas. *Lancet* **II:** 566 – 567

Appendix A – Primers (oligonucleotide sequences). The primers 5'cfpro, 3'cfpro, loxP, cf5491 and cmv1, which were used in vector construction, were HPLC-purified to decrease the possibility of aberrant sequences within these primers. All other primers were made by 3 OD select, desalted purification.

Name	Sequence (5' → 3')	Location in plasmids + orientation*
5'cfpro	GCTCCCCGCGGGCCCGGGCTCCACGCG TATTTGTTTAAACTATGGCGCGCCAAC TTTTCGGCTCTCTAAGGCTG	p1kbcfproEGFP: 69 → 136 (f) (excluding 1st 6 bp of oligo)
3'cfpro	CGCGGATCCTCGCTACCTTAGGACCGT TATAGTTAGGTCTCTCGGGCGCTGGGG	p1kbcfproEGFP: 867 → 814 (r)
loxP	GCTTACATAACTTCGTATAGCATACAT TATACGAAGTTATCATATGACAAACCA CAACTAGAATGC	p1kbcfproEGFP: 1865 → 1800 (r)
βgal1	GAATTATTTTTGATGGCG	pcmvβ:1363→1380 (f)
βgal2	CGCTGATTTGTGTAGTCGGTT	pcmvβ: 1614 → 1594 (r)
βgalint	TGATGGTGCTGCGTTGGAG	pcmvβ: 1501 → 1529 (f)
EGFPs	GCTGTTCACCGGGGTGGTGC	pEGFP-N 699 → 718 (f)
EGFPas	CATGATATAGCAGTTGTGGC	pEGFP-N 1140→1121 (r)
GAPDH 5'	TGGAGCCAAACGGGTCATCA	NA
GAPDH exon8	GTGGCAGTGATGGCATGGAC	NA
GAPDH int	TCCTGCACCACCAACTGCTTA	NA
Ex4F	GGATAACAAGGAGGAACGCTCTATCG	$_{PAC}$RC1b: 66414 → 66439 (f)
Ex9R	CCTGCTCCAGTGGATCCAGCAAC	$_{PAC}$RC1b: 67469 → 67447 (r)
Ex19F	GCGATCTGTGAGCCGAGTC	$_{PAC}$RC1b: 69552 → 69570 (f)
Ex24R	CCGCACTTTGTTCTCTTCTATG	$_{PAC}$RC1b: 70347 → 70326 (r)
seqF	GGACTCAGATCTCGAGCTCAAG	p1kbcfproEGFP 21→ 42 (f)
seqR	CCAGCTCGACCAGGATGGGCACC	p1kbcfproEGFP 935 → 913 (r)
seqF2	GGAAACGCCTGGTATCTTTATAGTCC	p1kbcfproEGFP 4720 → 4745 (f)
seqR2	GAAGAAGATGGTGCGCTCCTGGAC	p1kbcfproEGFP 1180 → 1157 (r)
pr1F	CCACCCTTGGAGTTCACTCACC	$_{PAC}$65kbcfproEGFP: 65253 → 65274 (f)
pr1R	TCCTTCCTCCTCTCCTCCTTCG	$_{PAC}$65kbcfproEGFP: 65791 → 65770 (r)

pr2F	AGGGAGGCTGGGAGTCAGAATC	PAC65kbcfproEGFP: 65718 → 65739 (f)
pr2R	CTTCAGGGTCAGCTTGCCGTAG	PAC65kbcfproEGFP: 66264 → 66243 (r) PACRC1iresEGFP → 71306 (r) PACRC2iresEGFP: → 95411 (r)
pr3F	ATCCTGGTCGAGCTGGACGGC	PAC65kbcfproEGFP: 66169 → 66198 (f) PACRC1iresEGFP 71232 → (f) PACRC2iresEGFP 95337 →
pr3R	CTGCTGGTAGTGGTCGGCGAG	PAC65kbcfproEGFP: 66681 → 66661 (r) PACRC1iresEGFP → 71724 (r) PACRC2iresEGFP: → 95829 (r)
pr4F	TCATGGCCGACAAGCAGAAGAAC	PAC65kbcfproEGFP: 66584 → 66606 (f) PACRC1iresEGFP: 71647 → (f) PACRC2iresEGFP: 95752 →
pr4R	TCCCATAATGGTGAAAGTTCCCTC	PAC65kbcfproEGFP: 67183 → 67160 (r) PACRC1iresEGFP: → 72223 PACRC2iresEGFP: → 96330
cf5420	TGCTGAACGAGAGGAGCCTC	PACRC1iresEGFP: 70370 → 70389 (f) PACRC2iresEGFP: 94475 →
cf5491	AACTCAAGCAAGTGCAAGTCTAAG	PACRC1iresEGFP: 70441 → 70464 (f) PACRC2iresEGFP: 94546→ 94569 (f)
cmvseq	GTCTCCACCCCATTGACGTC	pEGFP-N 429 → 448 PACRC2b
cmv1	GTAACAACTCCGCCCCATTGACGC	pEGFP-N 498 → 521 PAC RC2b

*for orientation, (f) = forward (r)=reverse

Appendix B: Sequencing results.

KEY:

e.g.

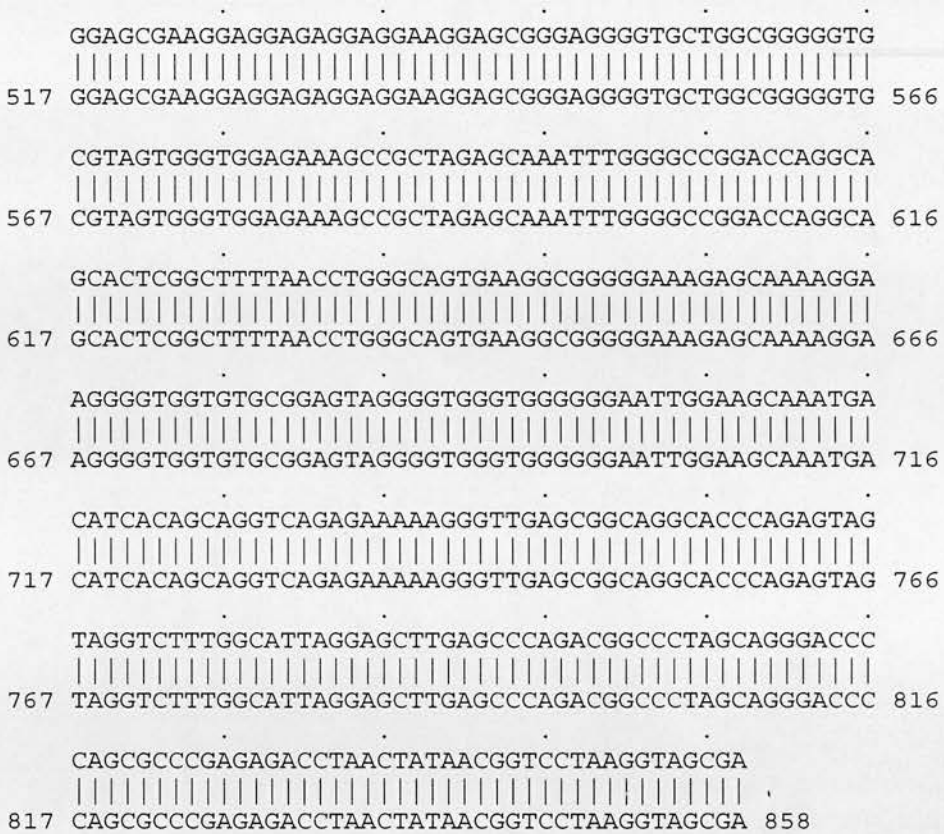
```
C                C                C
|  = match        = deletion       = mismatch
C                .                G
```

Top row = Contig, bottom row = Predicted sequence

B1: Sequencing results for p1kbcfproEGFP:

Percent Similarity: 100.000

```
                .         .         .         .         .
    CCGCGGGCCCGGGCTCCACGCGTATTTGTTTAAACTATGGCGCGCCAACT
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 67 CCGCGGGCCCGGGCTCCACGCGTATTTGTTTAAACTATGGCGCGCCAACT 116
                .         .         .         .         .
    TTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTT
    |||||||||||||||||||||||||||||||||||||||||||||||||||
117 TTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTT 166
                .         .         .         .         .
    CCCTTTTCAAAATGCACCTTGCAAACGTAACAGGAACCCGACTAGGATCA
    |||||||||||||||||||||||||||||||||||||||||||||||||||
167 CCCTTTTCAAAATGCACCTTGCAAACGTAACAGGAACCCGACTAGGATCA 216
                .         .         .         .         .
    TCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGC
    |||||||||||||||||||||||||||||||||||||||||||||||||||
217 TCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGC 266
                .         .         .         .         .
    AGCGGGTCCTGGGTCTGGCGGACCCTGACGCGAAGGAGGGTCTAGGAAGC
    |||||||||||||||||||||||||||||||||||||||||||||||||||
267 AGCGGGTCCTGGGTCTGGCGGACCCTGACGCGAAGGAGGGTCTAGGAAGC 316
                .         .         .         .         .
    TCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCAGCGT
    |||||||||||||||||||||||||||||||||||||||||||||||||||
317 TCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCAGCGT 366
                .         .         .         .         .
    TGCCAACTGGACCTAAAGAGAGGCCGCGACTGTCGCCCACCTGCGGGATG
    |||||||||||||||||||||||||||||||||||||||||||||||||||
367 TGCCAACTGGACCTAAAGAGAGGCCGCGACTGTCGCCCACCTGCGGGATG 416
                .         .         .         .         .
    GGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCGCGAG
    |||||||||||||||||||||||||||||||||||||||||||||||||||
417 GGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCGCGAG 466
                .         .         .         .         .
    GGAGGGAGGCTGGGAGTCAGAATCGGGAAAGGGAGGTGCGGGGCGGCGAG
    |||||||||||||||||||||||||||||||||||||||||||||||||||
467 GGAGGGAGGCTGGGAGTCAGAATCGGGAAAGGGAGGTGCGGGGCGGCGAG 516
```


GGAGCGAAGGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGGGGTG
517 GGAGCGAAGGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGGGGTG 566
CGTAGTGGGTGGAGAAAGCCGCTAGAGCAAATTTGGGGCCGGACCAGGCA
567 CGTAGTGGGTGGAGAAAGCCGCTAGAGCAAATTTGGGGCCGGACCAGGCA 616
GCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGGGGAAAGAGCAAAAGGA
617 GCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGGGGAAAGAGCAAAAGGA 666
AGGGGTGGTGTGCGGAGTAGGGGTGGGTGGGGGGAATTGGAAGCAAATGA
667 AGGGGTGGTGTGCGGAGTAGGGGTGGGTGGGGGGAATTGGAAGCAAATGA 716
CATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCAGAGTAG
717 CATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCAGAGTAG 766
TAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCC
767 TAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCC 816
CAGCGCCCGAGAGACCTAACTATAACGGTCCTAAGGTAGCGA
817 CAGCGCCCGAGAGACCTAACTATAACGGTCCTAAGGTAGCGA 858

B2: Sequence of $_{PAC}$65kbcfproEGFP:

Percent Similarity: 100.000

```
                .         .         .         .         .
        AACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65362   AACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTT 65411
                .         .         .         .         .
        CCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAGGAACCCGACTAGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65412   CCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAGGAACCCGACTAGG 65461
                .         .         .         .         .
        ATCATCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65462   ATCATCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGG 65511
                .         .         .         .         .
        TGGCAGCGGGTCCTGGGTCTGGCGGACCCTGACGCGAAGGAGGGTCTAGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65512   TGGCAGCGGGTCCTGGGTCTGGCGGACCCTGACGCGAAGGAGGGTCTAGG 65561
                .         .         .         .         .
        AAGCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65562   AAGCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCA 65611
                .         .         .         .         .
        GCGTTGCCAACTGGACCTAAAGAGAGGCCGCGACTGTCGCCCACCTGCGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65612   GCGTTGCCAACTGGACCTAAAGAGAGGCCGCGACTGTCGCCCACCTGCGG 65661
                .         .         .         .         .
        GATGGGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65662   GATGGGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCG 65711
                .         .         .         .         .
        CGAGGGAGGGAGGCTGGGAGTCAGAATCGGGAAAGGGAGGTGCGGGGCGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65712   CGAGGGAGGGAGGCTGGGAGTCAGAATCGGGAAAGGGAGGTGCGGGGCGG 65761
                .         .         .         .         .
        CGAGGGAGCGAAGGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65762   CGAGGGAGCGAAGGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGG 65811
                .         .         .         .         .
        GGTGCGTAGTGGGTGGAGAAAGCCGCTAGAGCAAATTTGGGGCCGGACCA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65812   GGTGCGTAGTGGGTGGAGAAAGCCGCTAGAGCAAATTTGGGGCCGGACCA 65861
                .         .         .         .         .
        GGCAGCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGGGGAAAGAGCAAA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65862   GGCAGCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGGGGAAAGAGCAAA 65911
                .         .         .         .         .
        AGGAAGGGGTGGTGTGCGGAGTAGGGGTGGGTGGGGGGAATTGGAAGCAA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65912   AGGAAGGGGTGGTGTGCGGAGTAGGGGTGGGTGGGGGGAATTGGAAGCAA 65961
                .         .         .         .         .
        ATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCAGA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
65962   ATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCAGA 66011
```

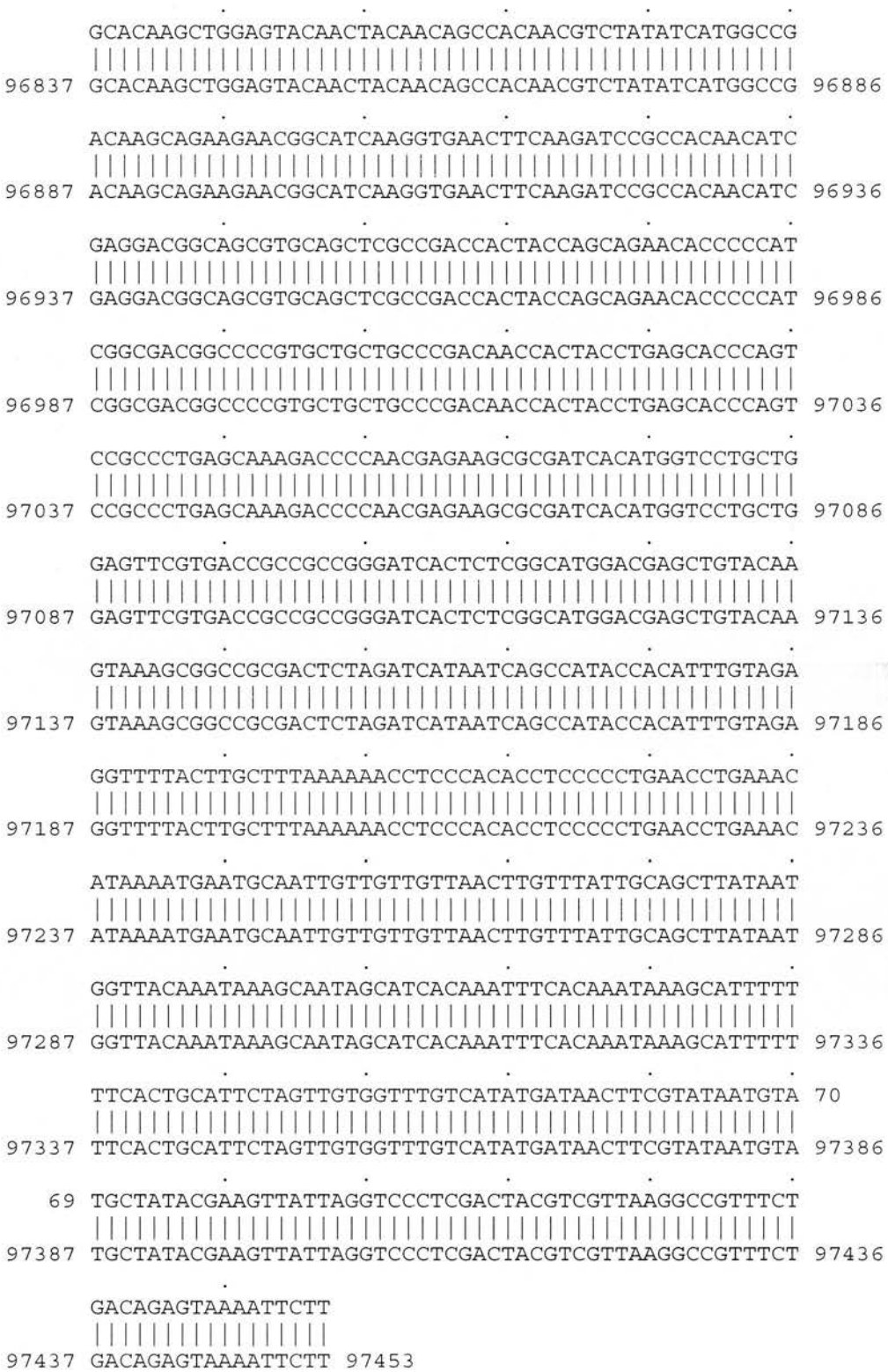
```
                .         .         .         .         .
      GTAGTAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGG
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66012 GTAGTAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGG 66061
                .         .         .         .         .
      ACCCCAGCGCCCGAGAGACCTAACTATAACGGTCCTAAGGTAGCGAGGAT
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66062 ACCCCAGCGCCCGAGAGACCTAACTATAACGGTCCTAAGGTAGCGAGGAT 66111
                .         .         .         .         .
      CCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66112 CCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT 66161
                .         .         .         .         .
      GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66162 GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA 66211
                .         .         .         .         .
      GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66212 GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG 66261
                .         .         .         .         .
      AAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66262 AAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT 66311
                .         .         .         .         .
      GACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66312 GACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA 66361
                .         .         .         .         .
      TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66362 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG 66411
                .         .         .         .         .
      GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66412 GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA 66461
                .         .         .         .         .
      GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66462 GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA 66511
                .         .         .         .         .
      TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66512 TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC 66561
                .         .         .         .         .
      TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCAT
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66562 TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCAT 66611
                .         .         .         .         .
      CAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGC
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66612 CAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGC 66661
                .         .         .         .         .
      TCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
      |||||||||||||||||||||||||||||||||||||||||||||||||||
66662 TCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG 66711
```

```
                .         .         .         .         .
        CTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
66712   CTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC 66761
                .         .         .         .         .
        CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
66762   CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG 66811
                .         .         .         .         .
        GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
66812   GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCT 66861
                .         .         .         .         .
        AGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
66862   AGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA 66911
                .         .         .         .         .
        AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
66912   AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGT 66961
                .         .         .         .         .
        TGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
66962   TGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATA 67011
                .         .         .         .         .
        GCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
67012   GCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGT 67061
                .         .         .         .
        GGTTTGTCATATGATAACTTCGTATAATGTATGCTATACGAAGTTAT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
67062   GGTTTGTCATATGATAACTTCGTATAATGTATGCTATACGAAGTTAT 67108
```

B3: Sequencing results for $_{PAC}$RC2cmvEGFP:

Percent Similarity: 100.000

```
                .         .         .         .         .
        CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96187   CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT 96236
                .         .         .         .         .
        GTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96237   GTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGG 96286
                .         .         .         .         .
        TGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96287   TGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTA 96336
                .         .         .         .         .
        GCGCTACCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGAC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96337   GCGCTACCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGAC 96386
                .         .         .         .         .
        GGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96387   GGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCG 96436
                .         .         .         .         .
        AGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96437   AGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC 96486
                .         .         .         .         .
        GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96487   GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC 96536
                .         .         .         .         .
          919   CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96537   CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG 96586
                .         .         .         .         .
        TGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96587   TGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC 96636
                .         .         .         .         .
        AGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96637   AGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT 96686
                .         .         .         .         .
        GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96687   GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA 96736
                .         .         .         .         .
        ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96737   ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC 96786
                .         .         .         .         .
        CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||
96787   CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG 96836
```

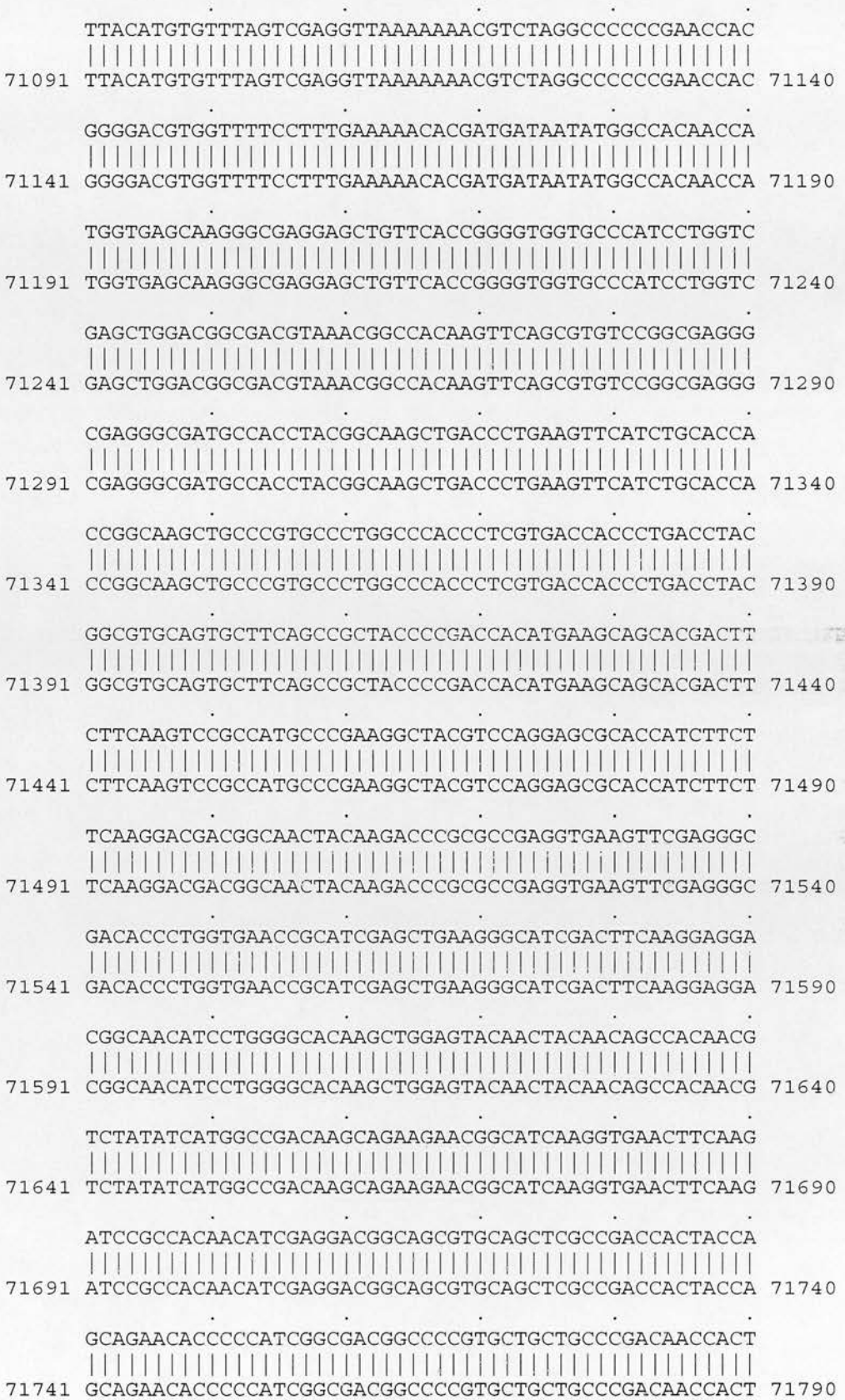

GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
96837 GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG 96886
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATC
96887 ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATC 96936
GAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT
96937 GAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT 96986
CGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
96987 CGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT 97036
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
97037 CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG 97086
GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAA
97087 GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAA 97136
GTAAAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGA
97137 GTAAAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGA 97186
GGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAAC
97187 GGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAAC 97236
ATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAAT
97237 ATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAAT 97286
GGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT
97287 GGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT 97336
TTCACTGCATTCTAGTTGTGGTTTGTCATATGATAACTTCGTATAATGTA 70
97337 TTCACTGCATTCTAGTTGTGGTTTGTCATATGATAACTTCGTATAATGTA 97386
69 TGCTATACGAAGTTATTAGGTCCCTCGACTACGTCGTTAAGGCCGTTTCT
97387 TGCTATACGAAGTTATTAGGTCCCTCGACTACGTCGTTAAGGCCGTTTCT 97436
GACAGAGTAAAATTCTT
97437 GACAGAGTAAAATTCTT 97453

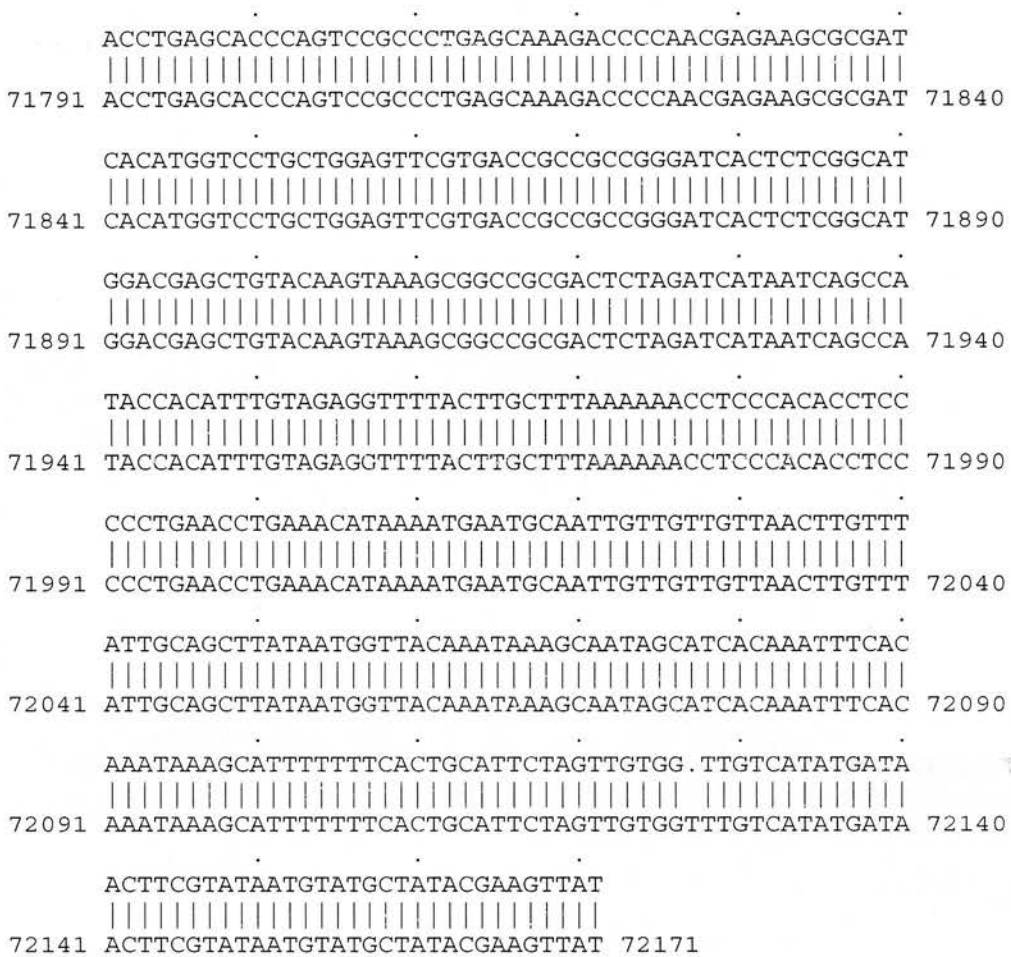
B4: Sequencing results for $_{PAC}$RC1iresEGFP:

Percent Similarity: 99.942

Mismatch at 71081, Deletion at 72127

```
                .         .         .         .         .
        AACTCAAGCAAGTGCAAGTCTAAGCCCCAGATTGCTGCTCTGAAAGAGGA
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70441   AACTCAAGCAAGTGCAAGTCTAAGCCCCAGATTGCTGCTCTGAAAGAGGA 70490
                .         .         .         .         .
        GACAGAAGAAGAGGTGCAAGATACAAGGCTTTAGAGAGCAGCATAAATGT
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70491   GACAGAAGAAGAGGTGCAAGATACAAGGCTTTAGAGAGCAGCATAAATGT 70540
                .         .         .         .         .
        TGACATGGGACATTTGCTCATGGAATTGGAGCTCGTCGACGGTACCGCGG
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70541   TGACATGGGACATTTGCTCATGGAATTGGAGCTCGTCGACGGTACCGCGG 70590
                .         .         .         .         .
        GCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAA
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70591   GCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAA 70640
                .         .         .         .         .
        GCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACC
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70641   GCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACC 70690
                .         .         .         .         .
        ATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTT
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70691   ATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTT 70740
                .         .         .         .         .
        CTTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAG
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70741   CTTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAG 70790
                .         .         .         .         .
        GTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGA
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70791   GTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGA 70840
                .         .         .         .         .
        CAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGG
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70841   CAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGG 70890
                .         .         .         .         .
        CGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAA
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70891   CGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAA 70940
                .         .         .         .         .
        AGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGA
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70941   AGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGA 70990
                .         .         .         .         .
  779   GTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCA 828
        |||||||||||||||||||||||||||||||||||||||||||||||||||
70991   GTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCA 71040
                .         .         .         .         .
        GAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTACACATGCT
        |||||||||||||||||||||||||||||||||||||||| ||||||||
71041   GAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCT 71090
```

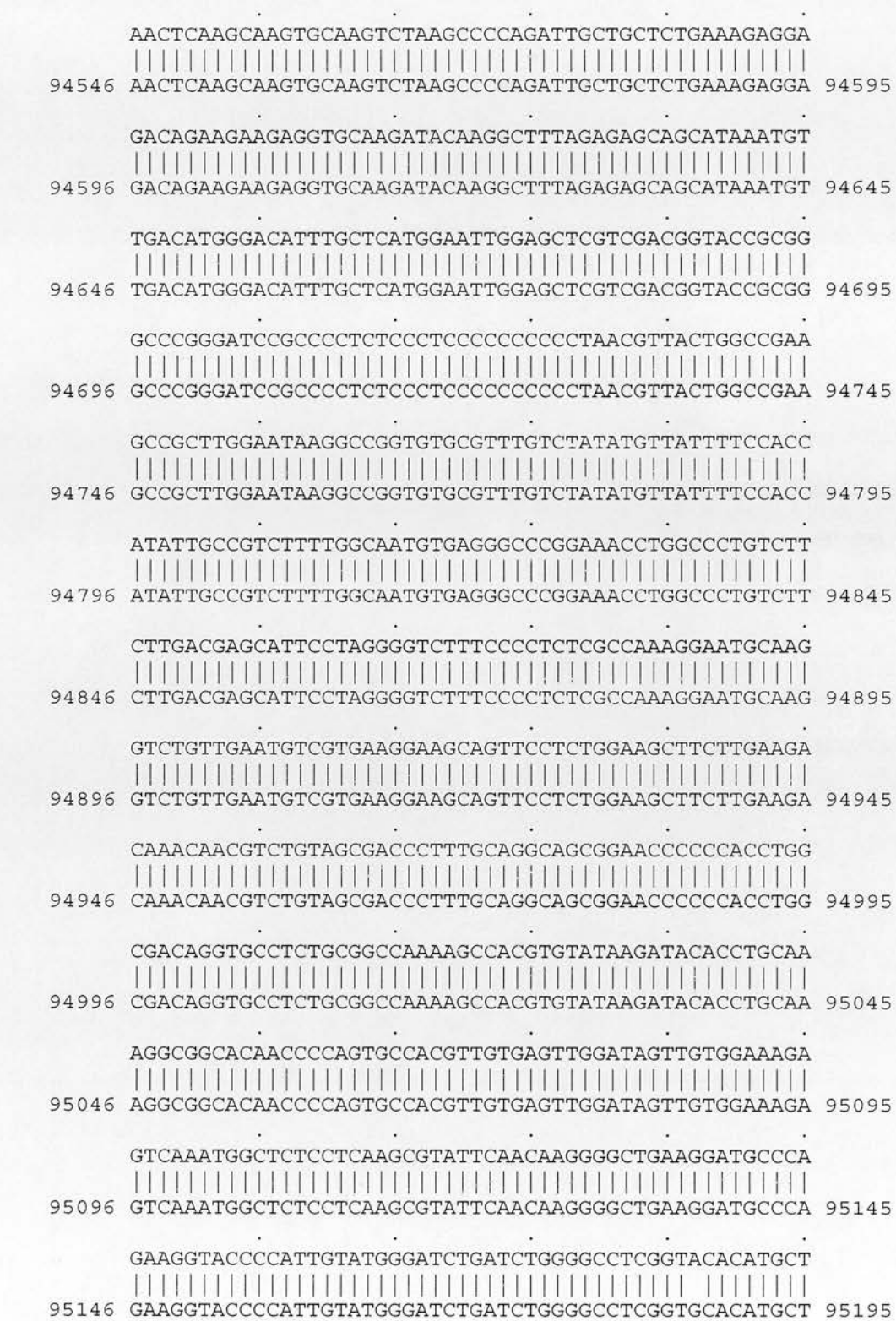

TTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCAC
71091 TTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCAC 71140
GGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGGCCACAACCA
71141 GGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGGCCACAACCA 71190
TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
71191 TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC 71240
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG
71241 GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG 71290
CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
71291 CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA 71340
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC
71341 CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC 71390
GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
71391 GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT 71440
CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
71441 CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT 71490
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
71491 TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC 71540
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
71541 GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA 71590
CGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACG
71591 CGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACG 71640
TCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAG
71641 TCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAG 71690
ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCA
71691 ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCA 71740
GCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
71741 GCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT 71790

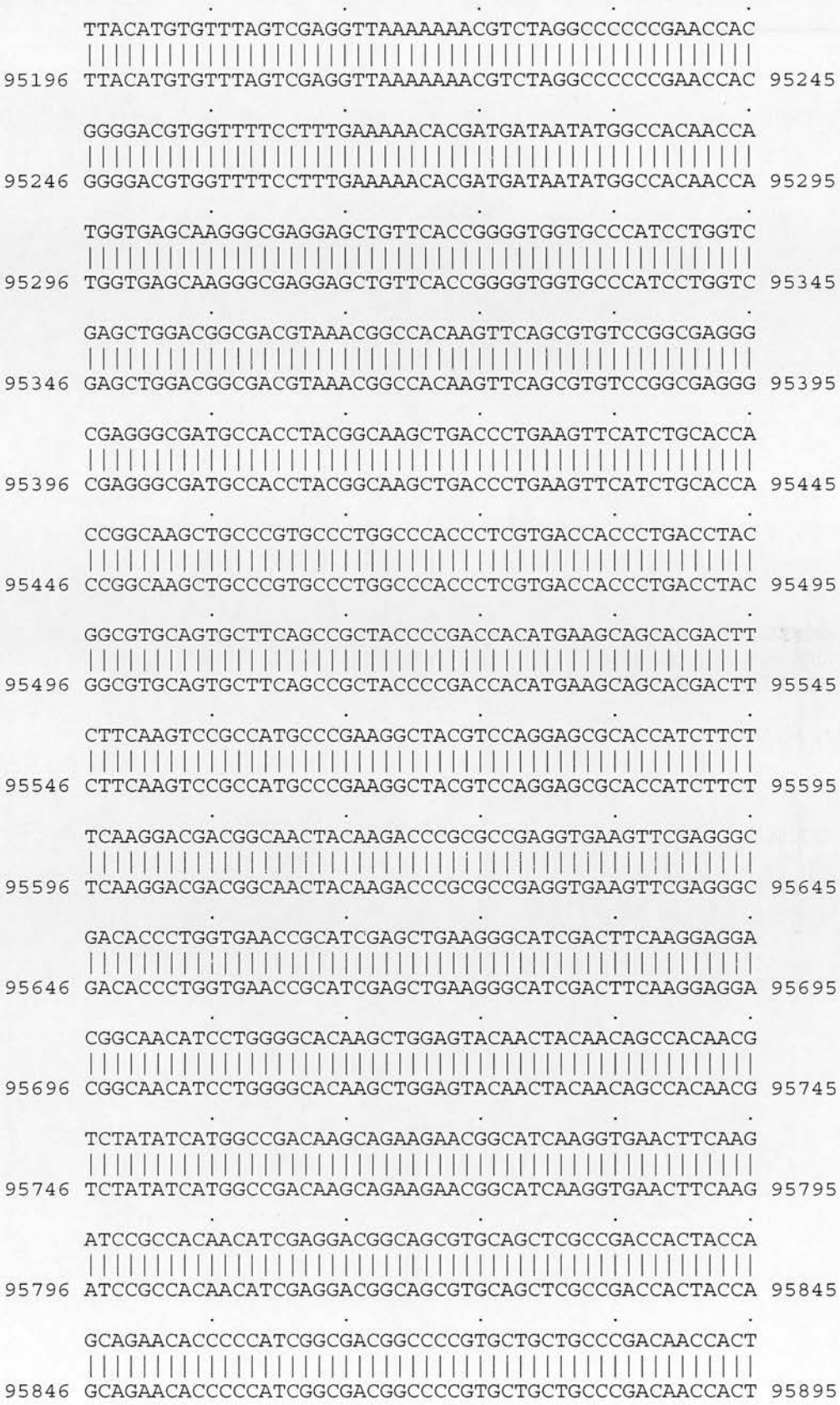

ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT
71791 ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT 71840
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
71841 CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT 71890
GGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCA
71891 GGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCA 71940
TACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCC
71941 TACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCC 71990
CCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTT
71991 CCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTT 72040
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC
72041 ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC 72090
AAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGG.TTGTCATATGATA
72091 AAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCATATGATA 72140
ACTTCGTATAATGTATGCTATACGAAGTTAT
72141 ACTTCGTATAATGTATGCTATACGAAGTTAT 72171

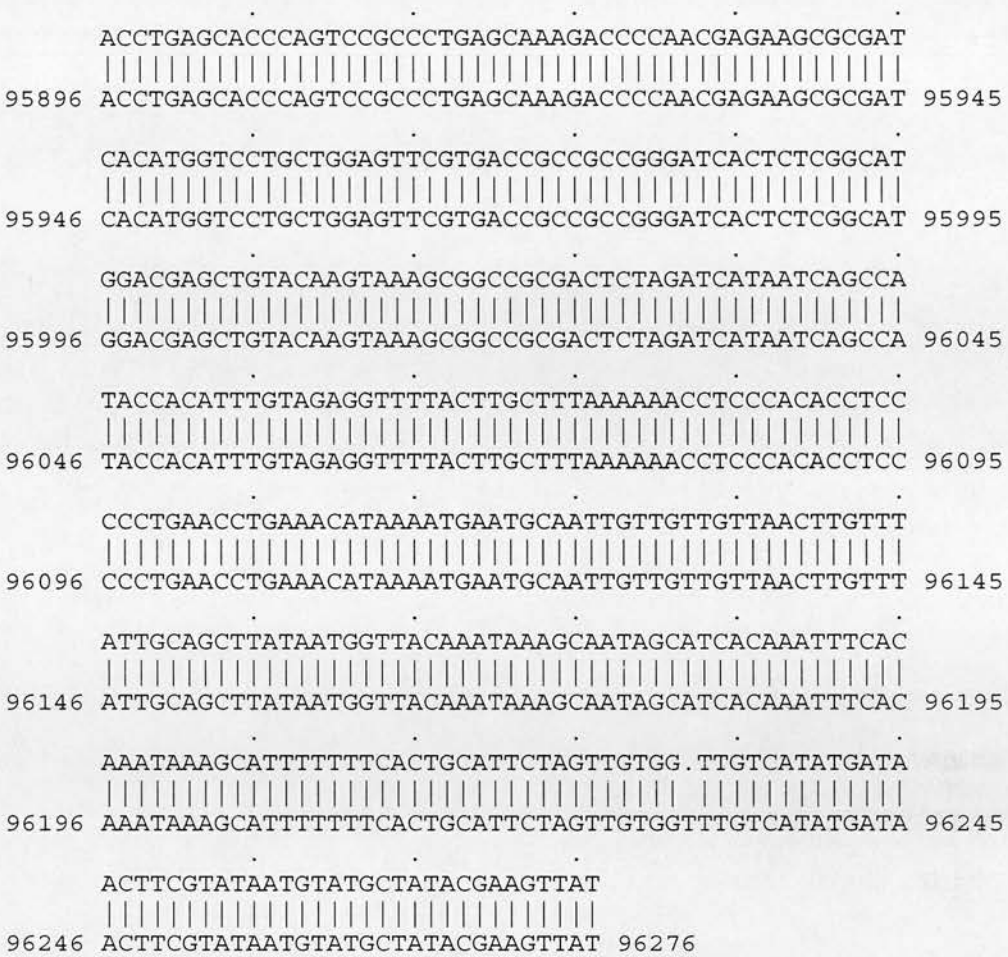
B5: Sequencing results for $_{PAC}$RC2iresEGFP:

Percent Similarity: 99.942

Mismatch at 95186, deletion at 96232



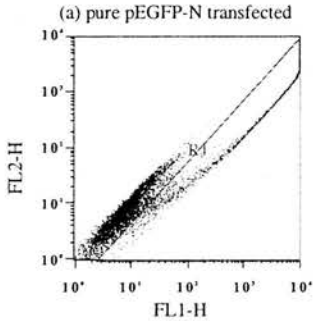

TTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCAC
95196 TTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCAC 95245
GGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGGCCACAACCA
95246 GGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGGCCACAACCA 95295
TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
95296 TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC 95345
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG
95346 GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG 95395
CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
95396 CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA 95445
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC
95446 CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC 95495
GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
95496 GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT 95545
CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
95546 CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT 95595
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
95596 TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC 95645
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
95646 GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA 95695
CGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACG
95696 CGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACG 95745
TCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAG
95746 TCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAG 95795
ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCA
95796 ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCA 95845
GCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
95846 GCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT 95895


ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT
95896 ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT 95945
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
95946 CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT 95995
GGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCA
95996 GGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCA 96045
TACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCC
96046 TACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCC 96095
CCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTT
96096 CCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTT 96145
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC
96146 ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC 96195
AAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGG.TTGTCATATGATA
96196 AAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCATATGATA 96245
ACTTCGTATAATGTATGCTATACGAAGTTAT
96246 ACTTCGTATAATGTATGCTATACGAAGTTAT 96276

Appendix C. Dotplot Data. This appendix shows the original FACS dotplots used to make the histogram figures.

C1 Figure 4.6 dotplots

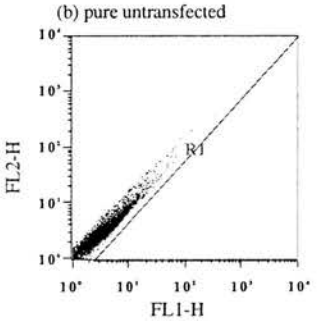
(a) pure pEGFP-N transfected



Statistics for: EGFPNU , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	613.32
	(X)				1829.35
R1	(Y)	3284	32.84	32.84	1715.44
	(X)				5416.26

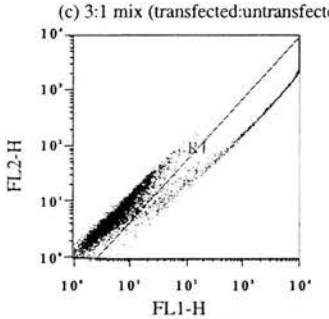
(b) pure untransfected



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.59
	(X)				3.86
R1	(Y)	0	0.00	0.00	--
	(X)				--

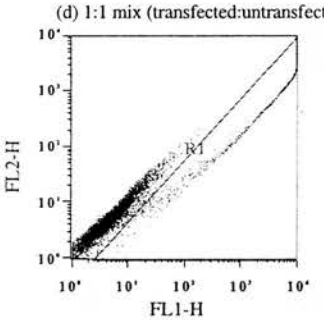
(c) 3:1 mix (transfected:untransfected)



Statistics for: EGFPN31 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	403.83
	(X)				1242.98
R1	(Y)	2217	22.17	22.17	1686.69
	(X)				5471.78

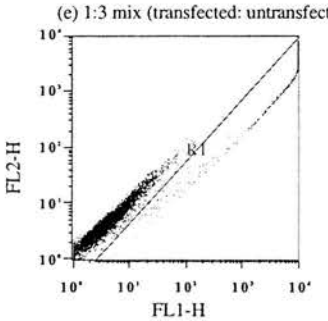
(d) 1:1 mix (transfected:untransfected)



Statistics for: EGFPN22 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	242.09
	(X)				729.52
R1	(Y)	1264	12.64	12.64	1720.43
	(X)				5577.83

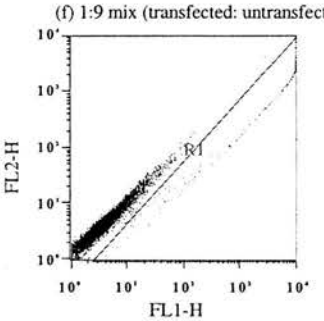
(e) 1:3 mix (transfected: untransfected)



Statistics for: EGFPN13 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	94.54
	(X)				275.27
R1	(Y)	464	4.64	4.64	1761.22
	(X)				5667.42

(f) 1:9 mix (transfected: untransfected)

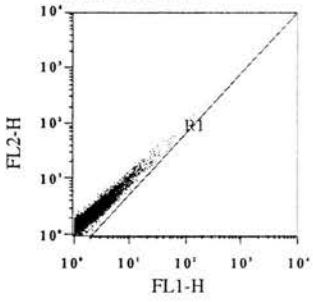


Statistics for: EGFPN19 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	35.54
	(X)				100.60
R1	(Y)	190	1.90	1.90	1403.48
	(X)				4857.00

C2 Figure 4.10 dotplots

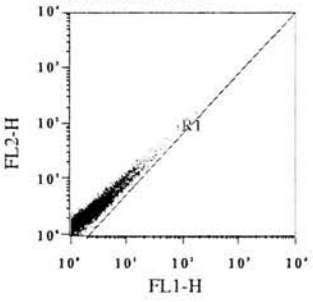
(a) untransfected -1



Statistics for: COSND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.20
	(X)				2.80
R1	(Y)	0	0.00	0.00	--
	(X)				--

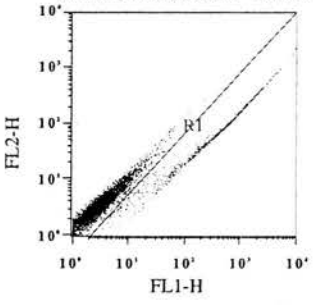
(b) untransfected -2



Statistics for: COSND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.16
	(X)				2.80
R1	(Y)	0	0.00	0.00	--
	(X)				--

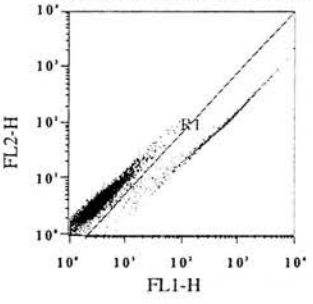
(c) PEI-p1kbcfproEGFP transfected -1



Statistics for: COSPEI1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	16.46
	(X)				78.66
R1	(Y)	1215	12.15	12.15	107.32
	(X)				622.69

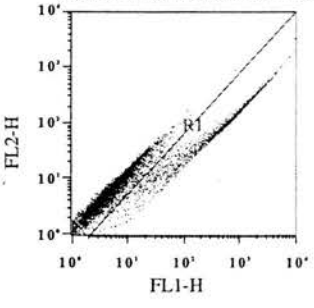
(d) PEI-p1kbcfproEGFP transfected -2



Statistics for: COSPEI2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	18.49
	(X)				89.60
R1	(Y)	1301	13.01	13.01	113.30
	(X)				663.42

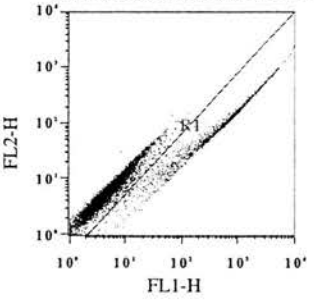
(e) Saint-Mix-p1kbcfproEGFP transfected -1



Statistics for: COSSNT1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	53.31
	(X)				277.49
R1	(Y)	2941	29.41	29.41	163.40
	(X)				929.16

(f) Saint-Mix-p1kbcfproEGFP transfected -2

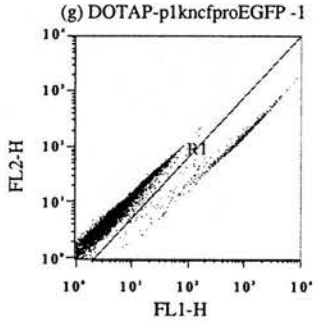


Statistics for: COSSNT2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	53.71
	(X)				278.49
R1	(Y)	2894	28.94	28.94	167.83
	(X)				947.96

C2 Figure 4.5 dotplots (continued)

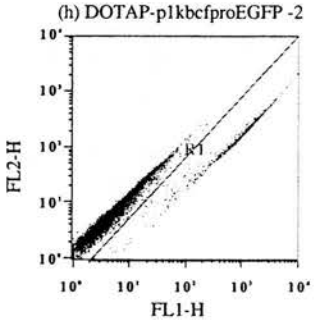
(g) DOTAP-p1kncfproEGFP -1



Statistics for: COSDTP1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	25.33
	(X)				106.96
R1	(Y)	921	9.21	9.21	195.47
	(X)				1100.96

(h) DOTAP-p1kbcfproEGFP -2

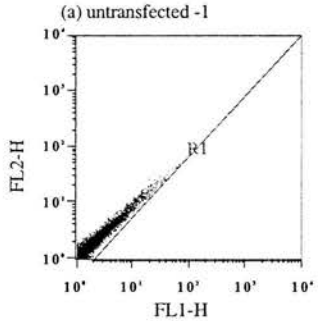


Statistics for: COSDTP2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	28.88
	(X)				122.90
R1	(Y)	971	9.71	9.71	217.44
	(X)				1205.20

C3 Figure 4.11 dotplots

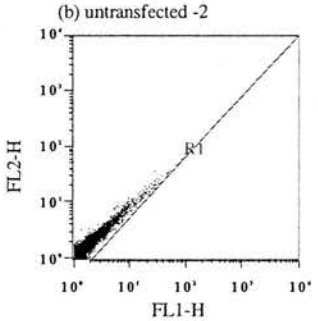
(a) untransfected -1



Statistics for: MDCKND1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	2.26
(X)				2.35
R1 (Y)	0	0.00	0.00	--
(X)				--

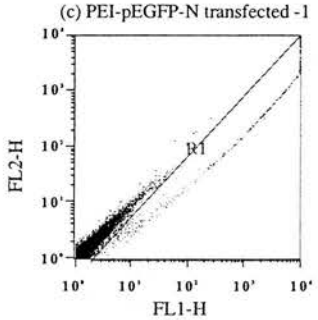
(b) untransfected -2



Statistics for: MDCKND2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	2.01
(X)				2.04
R1 (Y)	0	0.00	0.00	--
(X)				--

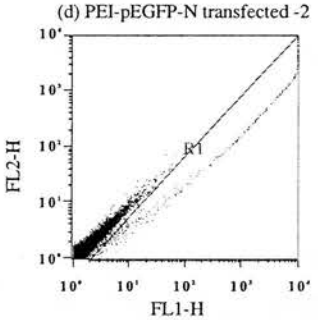
(c) PEI-pEGFP-N transfected -1



Statistics for: MDCKPEI1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	45.58
(X)				122.24
R1 (Y)	431	4.31	4.31	839.98
(X)				2615.34

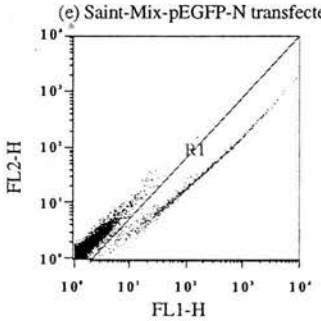
(d) PEI-pEGFP-N transfected -2



Statistics for: MDCKPEI2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	31.99
(X)				89.56
R1 (Y)	337	3.37	3.37	822.12
(X)				2525.83

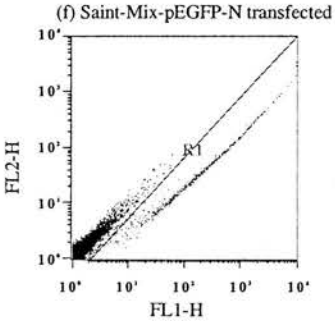
(e) Saint-Mix-pEGFP-N transfected -1



Statistics for: MDCKSNT1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	6.87
(X)				30.38
R1 (Y)	825	8.25	8.25	61.58
(X)				346.97

(f) Saint-Mix-pEGFP-N transfected -2

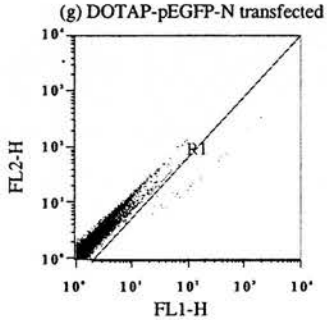


Statistics for: MDCKSNT2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	7.76
(X)				32.49
R1 (Y)	704	7.04	7.04	84.48
(X)				436.67

C3 Figure 4.11 dotplots (continued)

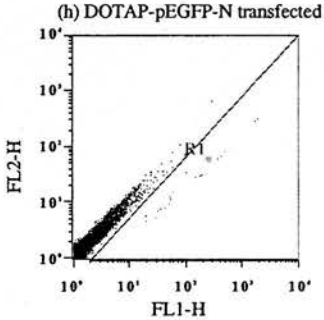
(g) DOTAP-pEGFP-N transfected -1



Statistics for: MDCKDTP1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.77
	(X)				3.04
R1	(Y)	26	0.26	0.26	47.28
	(X)				262.74

(h) DOTAP-pEGFP-N transfected -2

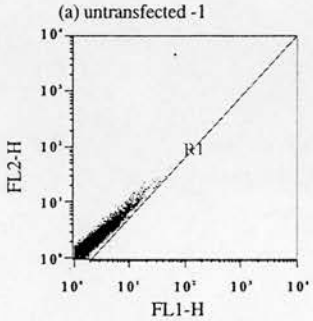


Statistics for: MDCKDTP2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.58
	(X)				2.89
R1	(Y)	25	0.25	0.25	49.58
	(X)				276.77

C4 Figure 4.15 dotplots

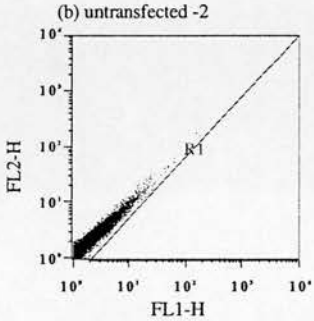
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.79
	(X)				3.06
R1	(Y)	0	0.00	0.00	--
	(X)				--

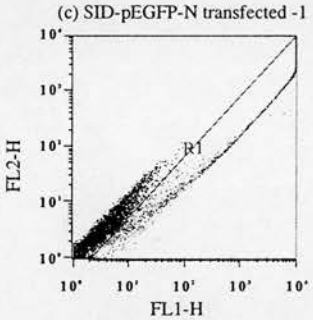
(b) untransfected -2



Statistics for: COSNDB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.72
	(X)				2.56
R1	(Y)	0	0.00	0.00	--
	(X)				--

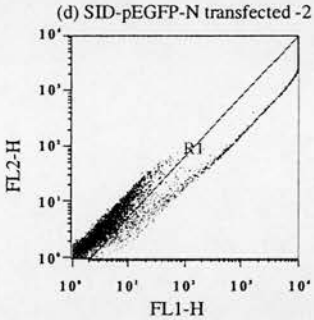
(c) SID-pEGFP-N transfected -1



Statistics for: COS1A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	189.63
	(X)				684.24
R1	(Y)	2005	20.05	20.05	914.40
	(X)				3381.49

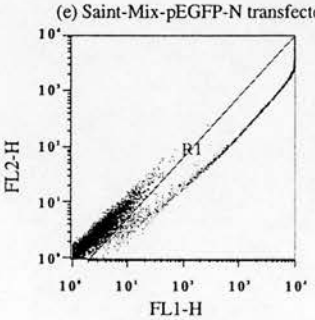
(d) SID-pEGFP-N transfected -2



Statistics for: COS1B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	212.12
	(X)				758.91
R1	(Y)	2008	20.08	20.08	994.96
	(X)				3718.83

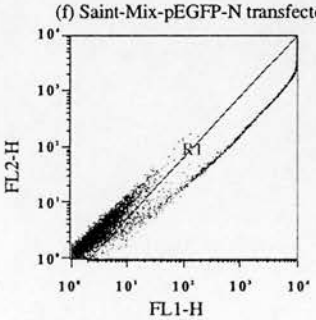
(e) Saint-Mix-pEGFP-N transfected -1



Statistics for: COS3A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	279.10
	(X)				941.19
R1	(Y)	2262	22.62	22.62	1200.15
	(X)				4127.41

(f) Saint-Mix-pEGFP-N transfected -2

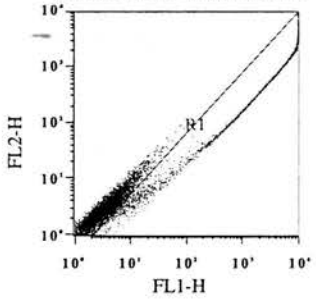


Statistics for: COS3B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	273.46
	(X)				931.77
R1	(Y)	2214	22.14	22.14	1177.86
	(X)				4151.59

C4 Figure 4.14 dotplots (continued)

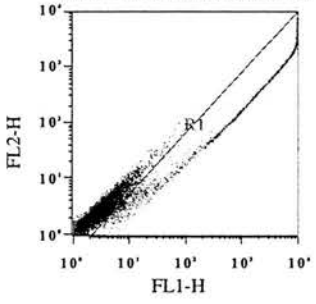
(g) LID-pEGFP-N transfected -1



Statistics for: COS4A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	356.81
	(X)				1302.67
R1	(Y)	3104	31.04	31.04	1126.23
	(X)				4172.72

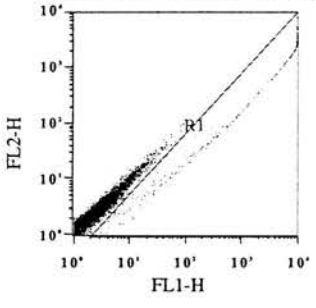
(h) LID-pEGFP-N transfected -2



Statistics for: COS4B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	306.29
	(X)				1070.09
R1	(Y)	2671	26.71	26.71	1123.15
	(X)				3982.07

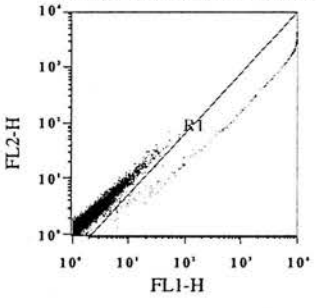
(i) Lipofectin -pEGFP-N transfected -1



Statistics for: COS5A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	34.60
	(X)				106.61
R1	(Y)	282	2.82	2.82	1117.08
	(X)				3674.55

(j) Lipofectin-pEGFP-N transfected -2

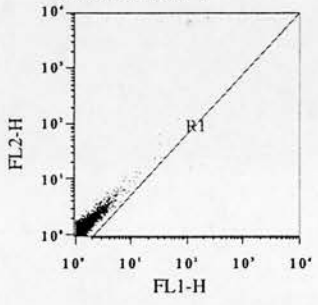


Statistics for: COS5B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	30.92
	(X)				104.05
R1	(Y)	285	2.85	2.85	985.50
	(X)				3557.37

C5 Figure 4.16 dotplots

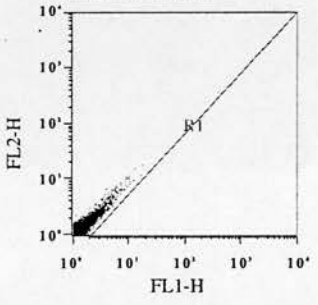
(a) untransfected -1



Statistics for: MDCKNDA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.37
	(X)				1.32
R1	(Y)	0	0.00	0.00	--
	(X)				--

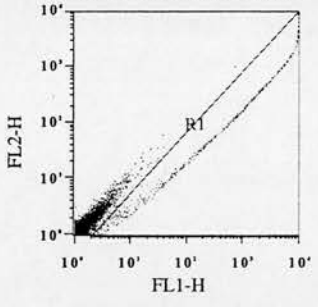
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.34
	(X)				1.38
R1	(Y)	0	0.00	0.00	--
	(X)				--

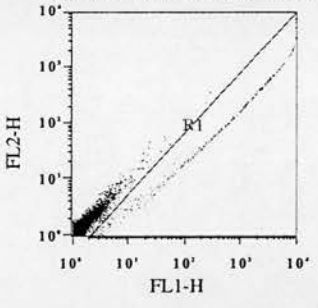
(c) SID-pEGFP-N transfected -1



Statistics for: MDCK2A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	31.86
	(X)				103.27
R1	(Y)	597	5.97	5.97	491.48
	(X)				1688.67

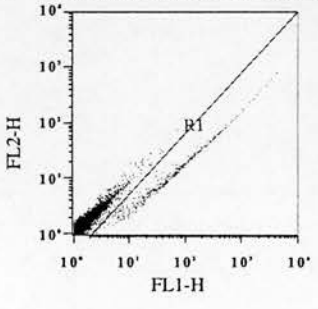
(d) SID-pEGFP-N transfected -2



Statistics for: MDCK2B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	17.69
	(X)				58.77
R1	(Y)	384	3.84	3.84	421.57
	(X)				1492.83

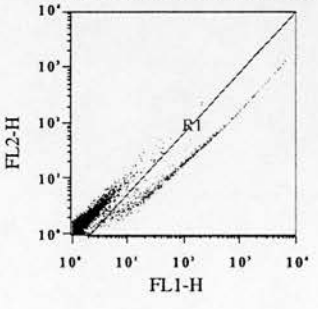
(e) Saint-Mix-pEGFP-N transfected -1



Statistics for: MDCK3A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.37
	(X)				11.44
R1	(Y)	504	5.04	5.04	36.33
	(X)				197.42

(f) Saint-Mix-pEGFP-N transfected -2

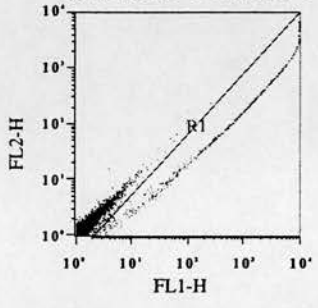


Statistics for: MDCK3B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.70
	(X)				17.56
R1	(Y)	687	6.87	6.87	44.58
	(X)				233.06

C5 Figure 4.16 dotplots (continued)

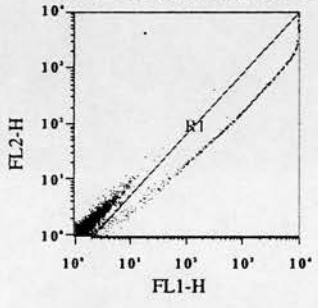
(g) LID-pEGFP-N transfected -1



Statistics for: MDCK4A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	40.87
	(X)				134.77
R1	(Y)	697	6.97	6.97	536.95
	(X)				1884.21

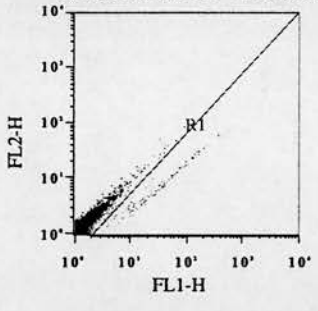
(h) LID-pEGFP-N transfected -2



Statistics for: MDCK4B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	41.47
	(X)				133.42
R1	(Y)	677	6.77	6.77	534.30
	(X)				1892.27

(i) Lipofectin-pEGFP-N transfected -1

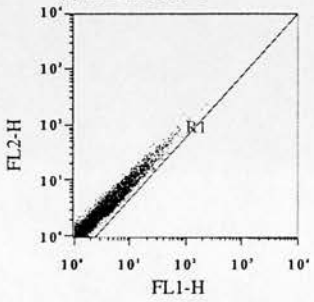


Statistics for: MDCK5A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.66
	(X)				2.40
R1	(Y)	137	1.37	1.37	13.58
	(X)				70.23

C6 Figure 4.17 dotplots

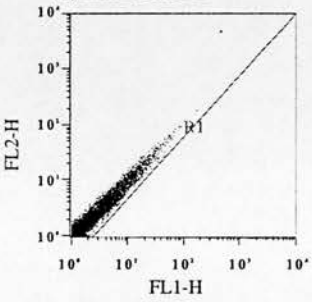
(a) untransfected -1



Statistics for: CACO2NDA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.28
	(X)				3.33
R1	(Y)	0	0.00	0.00	--
	(X)				--

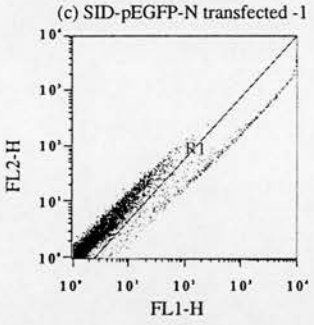
(b) untransfected -2



Statistics for: CACO2NDB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.10
	(X)				3.12
R1	(Y)	0	0.00	0.00	--
	(X)				--

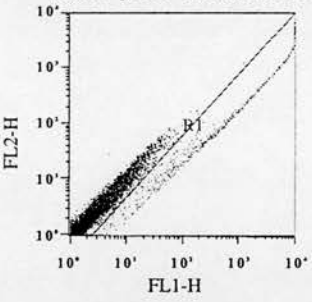
(c) SID-pEGFP-N transfected -1



Statistics for: CACO22A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	25.21
	(X)				84.87
R1	(Y)	601	6.01	6.01	364.99
	(X)				1359.96

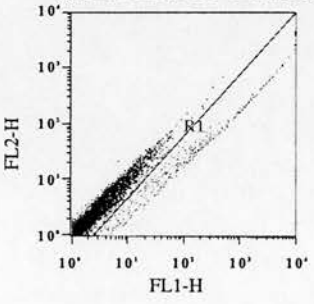
(d) SID-pEGFP-N transfected -2



Statistics for: CACO22B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	27.92
	(X)				93.45
R1	(Y)	620	6.20	6.20	393.66
	(X)				1452.38

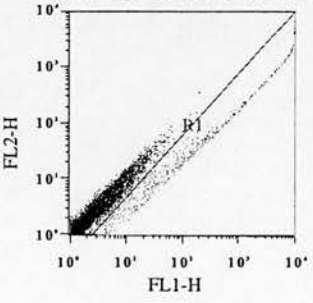
(e) Saint-Mix-pEGFP-N transfected -1



Statistics for: CACO23A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	12.33
	(X)				35.97
R1	(Y)	443	4.43	4.43	181.69
	(X)				720.20

(f) Saint-Mix-pEGFP-N transfected -2

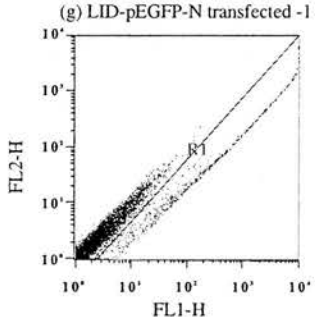


Statistics for: CACO23B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	14.87
	(X)				48.79
R1	(Y)	491	4.91	4.91	242.76
	(X)				935.47

C6 Figure 4.17 dotplots (continued)

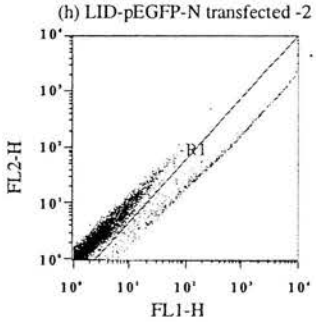
(g) LID-pEGFP-N transfected -1



Statistics for: CACO24A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	20.57
	(X)				78.22
R1	(Y)	809	8.09	8.09	224.85
	(X)				937.41

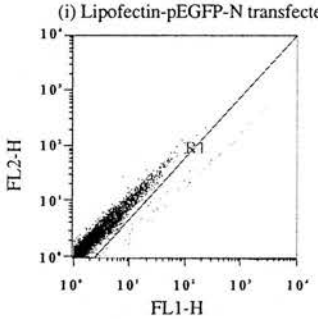
(h) LID-pEGFP-N transfected -2



Statistics for: CACO24B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	22.24
	(X)				77.39
R1	(Y)	797	7.97	7.97	246.53
	(X)				938.31

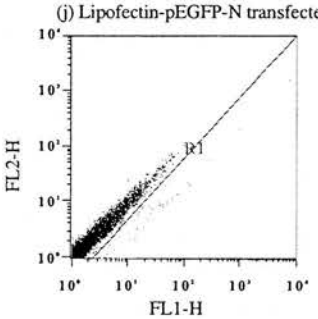
(i) Lipofectin-pEGFP-N transfected -1



Statistics for: CACO25A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.24
	(X)				5.00
R1	(Y)	56	0.56	0.56	78.82
	(X)				396.07

(j) Lipofectin-pEGFP-N transfected -2

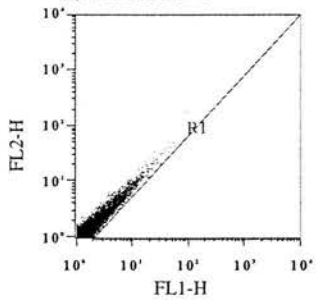


Statistics for: CACO25B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.96
	(X)				3.90
R1	(Y)	53	0.53	0.53	47.02
	(X)				215.14

C7 Figure 4.18 dotplots

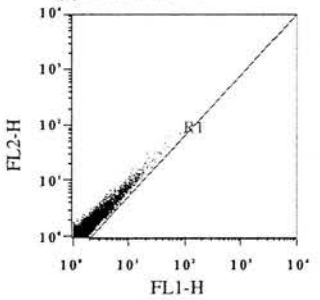
(a) untransfected -1



Statistics for: T84ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.09
	(X)				2.36
R1	(Y)	0	0.00	0.00	--
	(X)				--

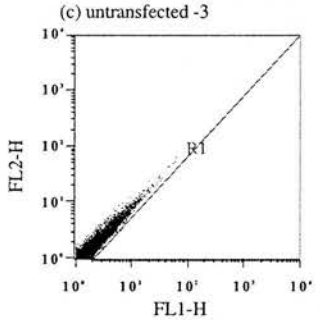
(b) untransfected -2



Statistics for: T84ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.05
	(X)				2.29
R1	(Y)	0	0.00	0.00	--
	(X)				--

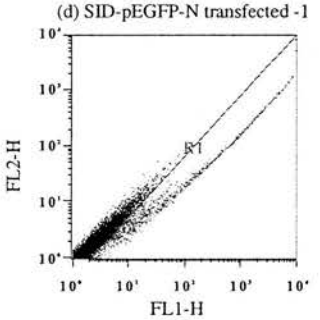
(c) untransfected -3



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.05
	(X)				2.32
R1	(Y)	0	0.00	0.00	--
	(X)				--

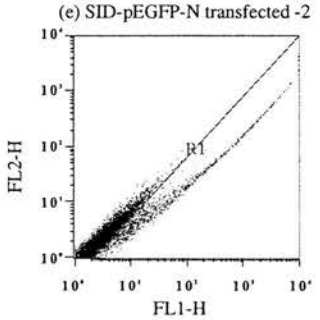
(d) SID-pEGFP-N transfected -1



Statistics for: T84SNTP61 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	14.68
	(X)				61.83
R1	(Y)	1423	14.23	14.23	84.43
	(X)				411.18

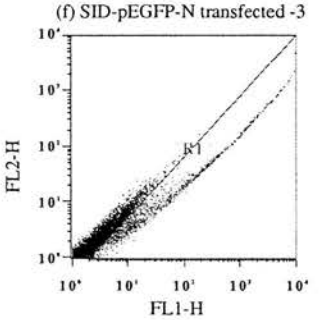
(e) SID-pEGFP-N transfected -2



Statistics for: T84SNTP62 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	14.02
	(X)				57.11
R1	(Y)	1623	16.23	16.23	71.55
	(X)				332.92

(f) SID-pEGFP-N transfected -3

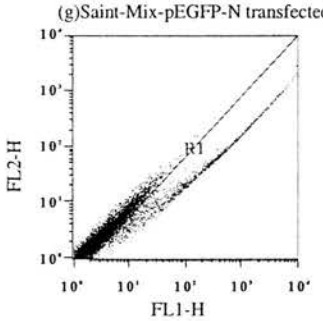


Statistics for: T84SNTP63 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	14.63
	(X)				64.75
R1	(Y)	1640	16.40	16.40	73.64
	(X)				375.37

C7 Figure 4.18 dotplots (continued)

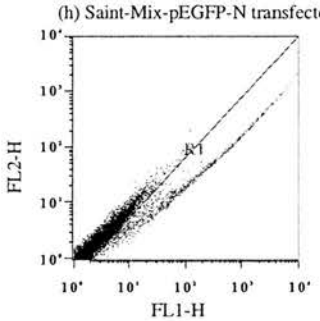
(g)Saint-Mix-pEGFP-N transfected -1



Statistics for: T84SNT2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	17.36
	(X)				80.24
R1	(Y)	1562	15.62	15.62	91.62
	(X)				490.01

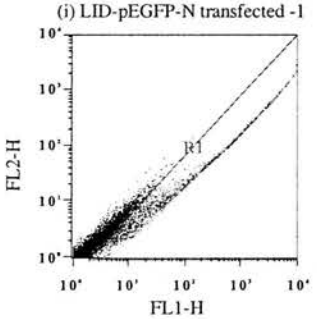
(h) Saint-Mix-pEGFP-N transfected -2



Statistics for: T84SNT3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	12.35
	(X)				52.10
R1	(Y)	1247	12.47	12.47	77.32
	(X)				391.29

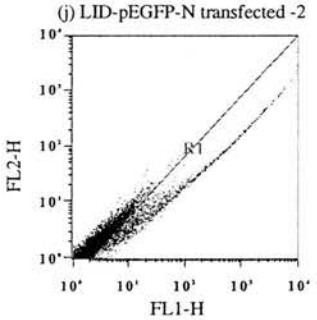
(i) LID-pEGFP-N transfected -1



Statistics for: T84LID1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	19.71
	(X)				92.59
R1	(Y)	2249	22.49	22.49	78.13
	(X)				399.27

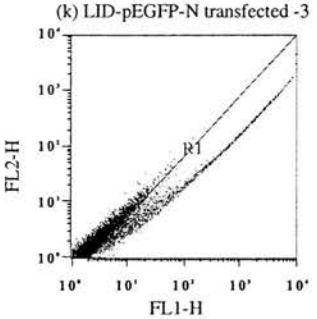
(j) LID-pEGFP-N transfected -2



Statistics for: T84LID2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	13.56
	(X)				61.67
R1	(Y)	2053	20.53	20.53	56.34
	(X)				287.55

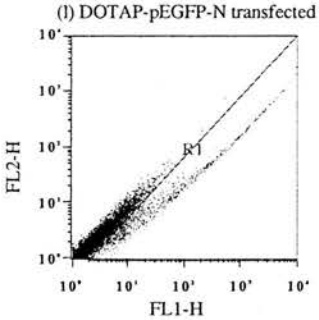
(k) LID-pEGFP-N transfected -3



Statistics for: T84LID3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	19.29
	(X)				91.33
R1	(Y)	2165	21.65	21.65	78.91
	(X)				408.48

(l) DOTAP-pEGFP-N transfected -1

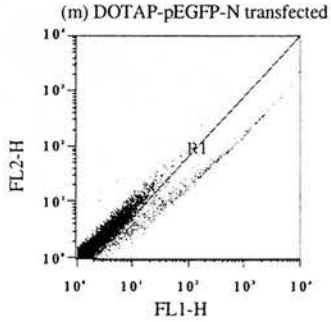


Statistics for: T84DTP1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.95
	(X)				29.69
R1	(Y)	933	9.33	9.33	54.23
	(X)				282.25

C7 Figure 4.18 dotplots (continued)

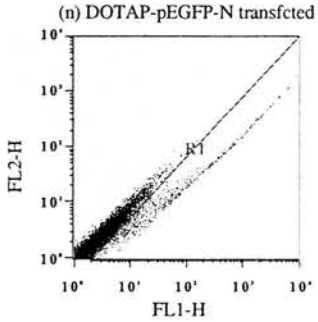
(m) DOTAP-pEGFP-N transfected -2



Statistics for: T84DTP2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.15
	(X)				29.99
R1	(Y)	811	8.11	8.11	66.55
	(X)				330.72

(n) DOTAP-pEGFP-N transfcted -3

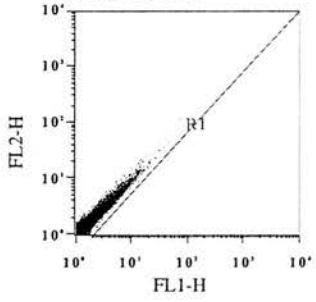


Statistics for: T84DTP3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.17
	(X)				34.44
R1	(Y)	813	8.13	8.13	76.47
	(X)				381.93

C8 Figure 4.19 dotplots

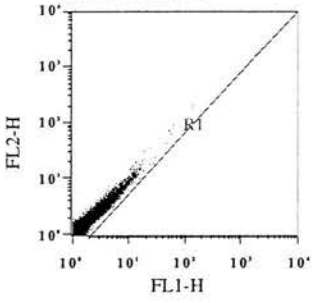
(a) untransfected -1



Statistics for: HBENDA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.40
	(X)				2.43
R1	(Y)	0	0.00	0.00	--
	(X)				--

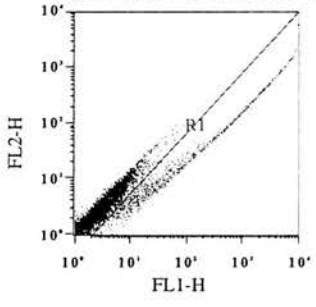
(b) untransfected -2



Statistics for: HBENDB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.52
	(X)				2.55
R1	(Y)	0	0.00	0.00	--
	(X)				--

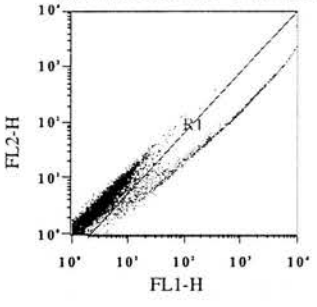
(c) SID-pEGFP-N transfection -1



Statistics for: HBESNT1A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	17.60
	(X)				76.62
R1	(Y)	1114	11.14	11.14	130.38
	(X)				661.02

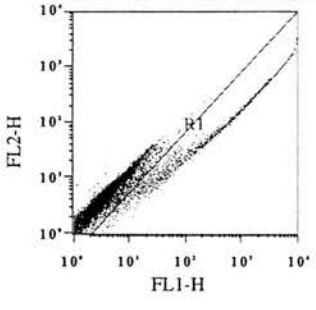
(d) Saint-Mix-pEGFP-N transfection -2



Statistics for: HBESNT1B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	24.29
	(X)				109.41
R1	(Y)	1186	11.86	11.86	173.89
	(X)				893.74

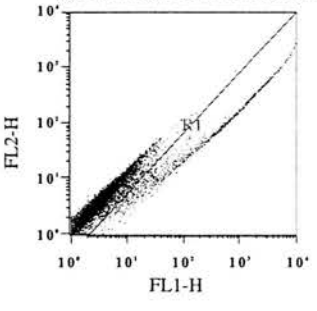
(e) Saint-Mix -pEGFP-N transfection -1



Statistics for: HBESNTO1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	31.71
	(X)				140.41
R1	(Y)	1431	14.31	14.31	189.43
	(X)				954.66

(f) Saint-Mix-pEGFP-N transfection -2

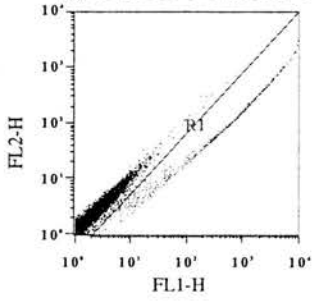


Statistics for: HBESNTO2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	24.44
	(X)				105.74
R1	(Y)	1215	12.15	12.15	164.64
	(X)				839.13

C8 Figure 4.19 dotplots (continued)

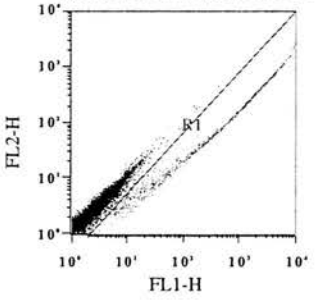
(g) LID-pEGFP-N transfected -1



Statistics for: HBEL4A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	17.30
	(X)				75.36
R1	(Y)	722	7.22	7.22	190.48
	(X)				999.26

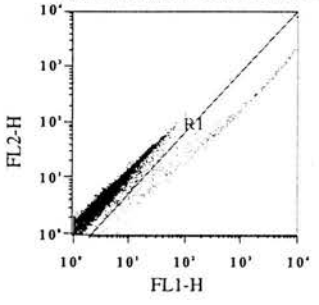
(h) LID-pEGFP-N transfected -2



Statistics for: HBEL4B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	19.50
	(X)				88.92
R1	(Y)	824	8.24	8.24	196.91
	(X)				1043.98

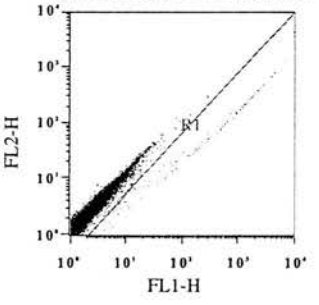
(i) DOTAP-pEGFP-N transfected -1



Statistics for: HBEDTPA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	13.51
	(X)				37.88
R1	(Y)	338	3.38	3.38	214.27
	(X)				972.45

(j) DOTAP-pEGFP-N transfected -2

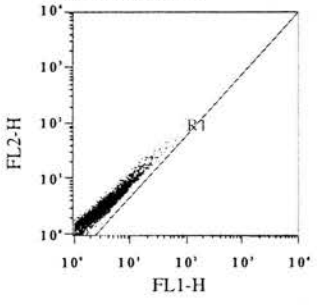


Statistics for: HBEDTPB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.03
	(X)				16.76
R1	(Y)	200	2.00	2.00	151.07
	(X)				675.71

C9 Figure 5.1 dotplots

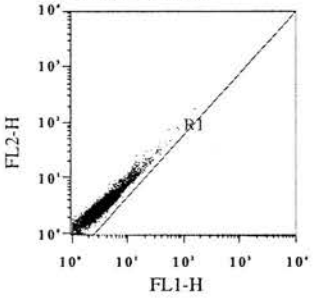
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.71
	(X)				3.96
R1	(Y)	0	0.00	0.00	--
	(X)				--

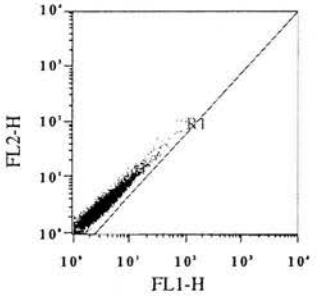
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.76
	(X)				4.05
R1	(Y)	0	0.00	0.00	--
	(X)				--

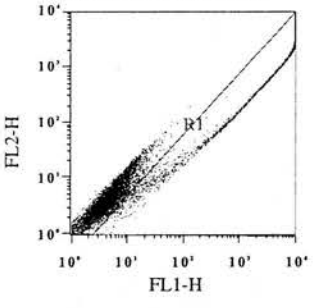
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.73
	(X)				4.01
R1	(Y)	0	0.00	0.00	--
	(X)				--

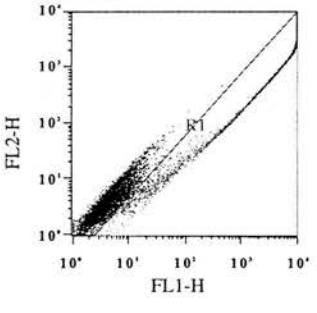
(d) pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	313.24
	(X)				1144.10
R1	(Y)	2792	27.92	27.92	1079.85
	(X)				4055.75

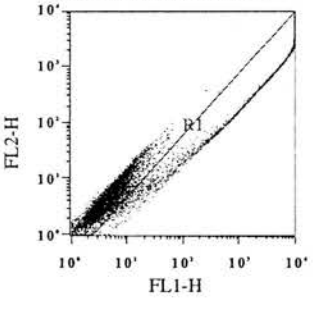
(e) pEGFP-N transfected -2



Statistics for: EGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	451.99
	(X)				1605.92
R1	(Y)	3684	36.84	36.84	1172.11
	(X)				4304.11

(f) pEGFP-N transfected -3

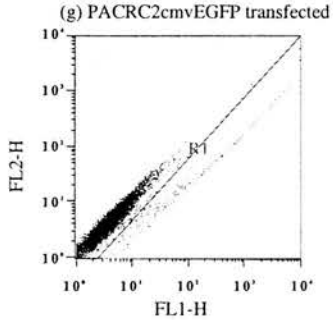


Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	443.72
	(X)				1608.91
R1	(Y)	3633	36.33	36.33	1175.93
	(X)				4383.70

C9 Figure 5.1 dotplots (continued)

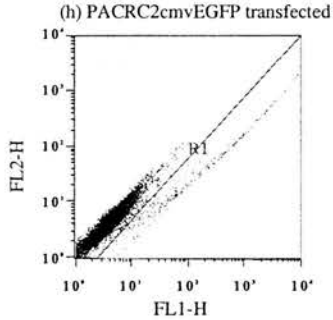
(g) PACRC2cmvEGFP transfected -1



Statistics for: RC2CG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.07
	(X)				16.83
R1	(Y)	289	2.89	2.89	106.42
	(X)				432.07

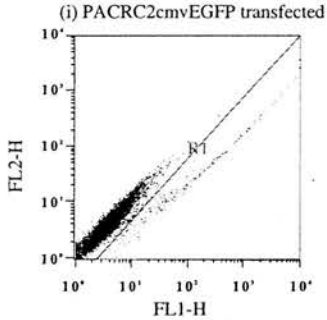
(h) PACRC2cmvEGFP transfected -2



Statistics for: RC2CG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.69
	(X)				24.99
R1	(Y)	390	3.90	3.90	128.87
	(X)				533.89

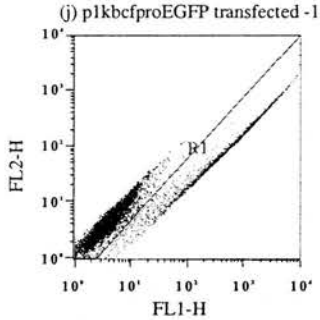
(i) PACRC2cmvEGFP transfected -3



Statistics for: RC2CG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	10.14
	(X)				26.21
R1	(Y)	365	3.65	3.65	150.71
	(X)				601.91

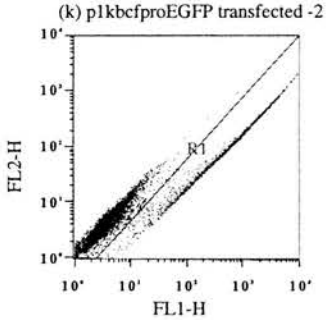
(j) p1kbcfproEGFP transfected -1



Statistics for: 1KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	69.35
	(X)				367.98
R1	(Y)	3441	34.41	34.41	191.78
	(X)				1060.89

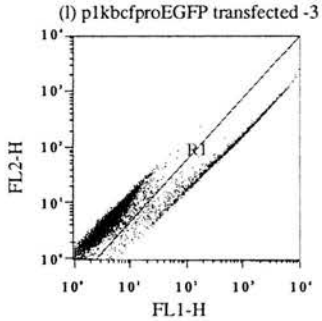
(k) p1kbcfproEGFP transfected -2



Statistics for: 1KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	81.74
	(X)				443.85
R1	(Y)	3979	39.79	39.79	197.37
	(X)				1108.68

(l) p1kbcfproEGFP transfected -3

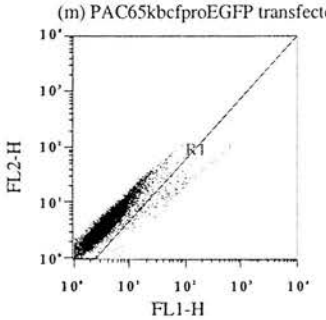


Statistics for: 1KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	78.53
	(X)				421.34
R1	(Y)	3798	37.98	37.98	198.13
	(X)				1102.00

C9 Figure 5.1 dotplots (continued)

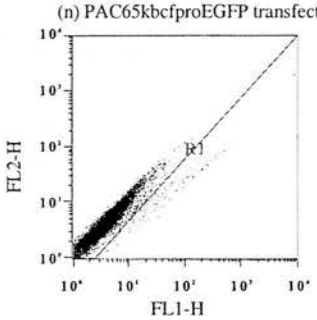
(m) PAC65kbcfproEGFP transfected -1



Statistics for: 65KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.68
	(X)				6.53
R1	(Y)	228	2.28	2.28	23.43
	(X)				92.56

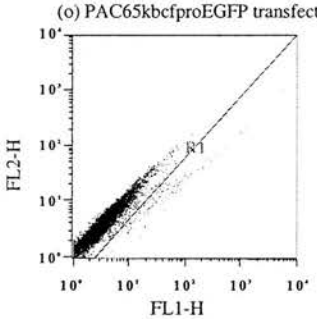
(n) PAC65kbcfproEGFP transfected -2



Statistics for: 65KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.74
	(X)				6.76
R1	(Y)	186	1.86	1.86	30.22
	(X)				125.03

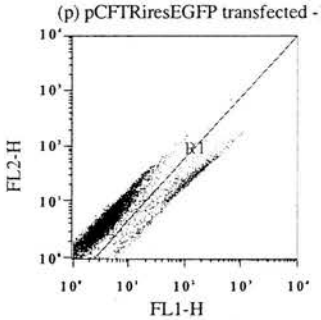
(o) PAC65kbcfproEGFP transfected -3



Statistics for: 65KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.38
	(X)				5.79
R1	(Y)	145	1.45	1.45	26.29
	(X)				112.66

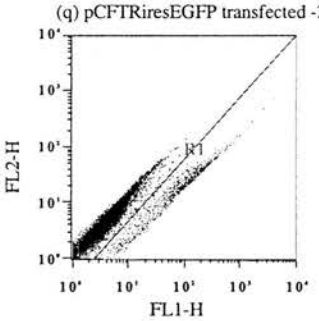
(p) pCFTRiresEGFP transfected -1



Statistics for: I2ES11 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.05
	(X)				18.21
R1	(Y)	1151	11.51	11.51	26.04
	(X)				121.30

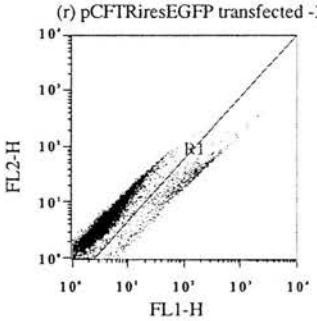
(q) pCFTRiresEGFP transfected -2



Statistics for: I2ES12 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.03
	(X)				17.51
R1	(Y)	990	9.90	9.90	28.18
	(X)				134.03

(r) pCFTRiresEGFP transfected -3

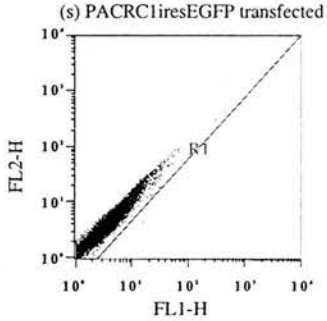


Statistics for: I2ES13 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.45
	(X)				16.46
R1	(Y)	857	8.57	8.57	30.65
	(X)				139.93

C9 Figure 5.1 dotplots (continued)

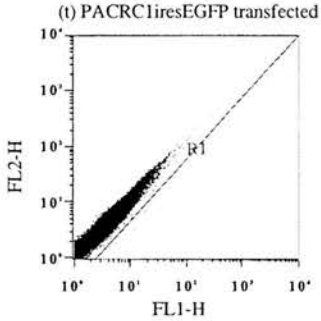
(s) PACRC1iresEGFP transfected -1



Statistics for: RC1IG1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	5.41
(X)				4.46
R1 (Y)	2	0.02	0.02	6.22
(X)				17.09

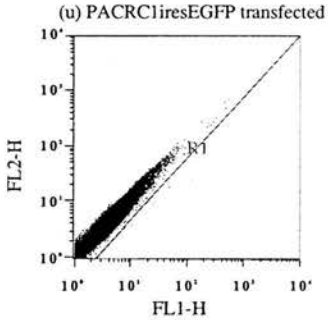
(t) PACRC1iresEGFP transfected -2



Statistics for: RC1IG2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	50000	100.00	100.00	5.09
(X)				4.19
R1 (Y)	9	0.02	0.02	11.09
(X)				39.46

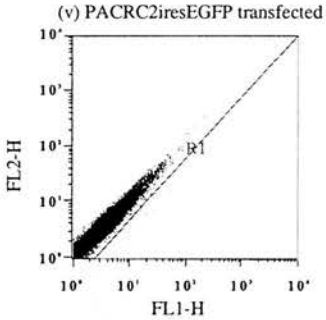
(u) PACRC1iresEGFP transfected -3



Statistics for: RC1IG3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	50000	100.00	100.00	5.59
(X)				4.76
R1 (Y)	13	0.03	0.03	15.91
(X)				44.24

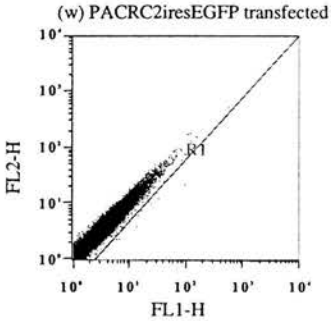
(v) PACRC2iresEGFP transfected -1



Statistics for: RC2IG2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	50000	100.00	100.00	4.50
(X)				4.10
R1 (Y)	21	0.04	0.04	7.61
(X)				20.61

(w) PACRC2iresEGFP transfected -2

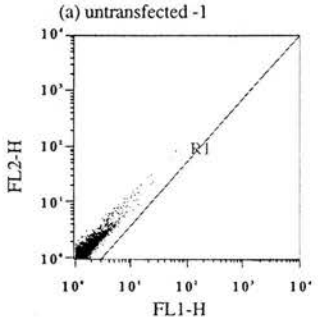


Statistics for: RC2IG3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	50000	100.00	100.00	4.44
(X)				4.09
R1 (Y)	15	0.03	0.03	8.41
(X)				23.55

C10 Figure 5.2 dotplots

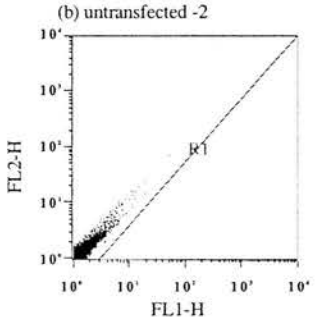
(a) untransfected -1



Statistics for: MND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.40
	(X)				1.40
R1	(Y)	0	0.00	0.00	--
	(X)				--

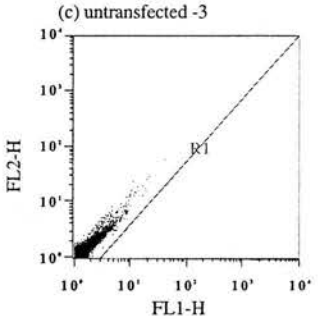
(b) untransfected -2



Statistics for: MND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.38
	(X)				1.39
R1	(Y)	0	0.00	0.00	--
	(X)				--

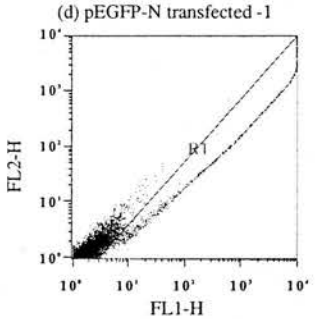
(c) untransfected -3



Statistics for: MND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.47
	(X)				1.51
R1	(Y)	0	0.00	0.00	--
	(X)				--

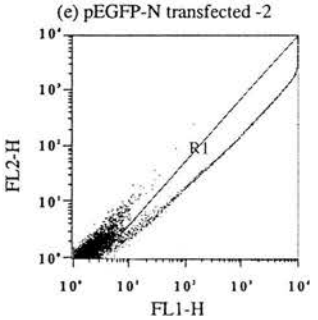
(d) pEGFP-N transfected -1



Statistics for: MEGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	83.39
	(X)				251.44
R1	(Y)	1205	12.05	12.05	638.46
	(X)				2028.80

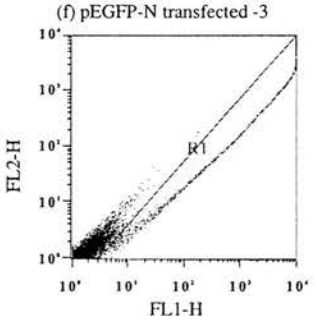
(e) pEGFP-N transfected -2



Statistics for: MEGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	92.87
	(X)				258.08
R1	(Y)	1202	12.02	12.02	695.03
	(X)				2065.17

(f) pEGFP-N transfected -3

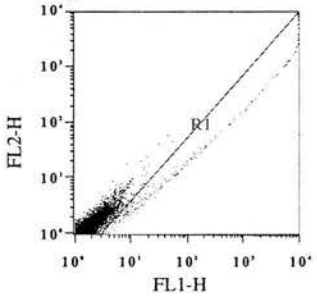


Statistics for: MEGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	75.55
	(X)				221.16
R1	(Y)	1101	11.01	11.01	655.58
	(X)				1974.18

C10 Figure 5.2 dotplots (continued)

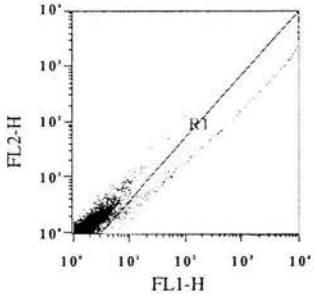
(g) PACRC2cmvEGFP transfected -1



Statistics for: MRC2CG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	28.27
	(X)				58.94
R1	(Y)	293	2.93	2.93	744.28
	(X)				1777.86

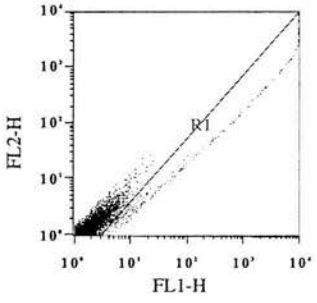
(h) PACRC2cmvEGFP transfected -2



Statistics for: MRC2CG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	25.06
	(X)				52.55
R1	(Y)	320	3.20	3.20	581.82
	(X)				1428.83

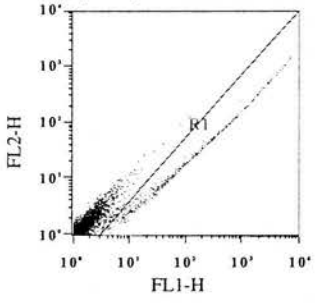
(i) PACRC2cmvEGFP transfected -3



Statistics for: MRC2CG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	25.26
	(X)				55.55
R1	(Y)	313	3.13	3.13	602.44
	(X)				1556.64

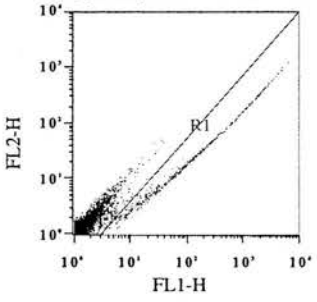
(j) p1kbcfproEGFP transfected -1



Statistics for: M1KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.65
	(X)				25.12
R1	(Y)	593	5.93	5.93	72.90
	(X)				400.67

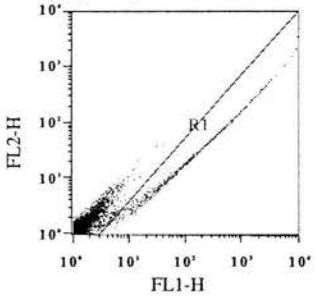
(k) p1kbcfproEGFP transfected -2



Statistics for: M1KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.35
	(X)				29.23
R1	(Y)	737	7.37	7.37	67.42
	(X)				377.82

(l) p1kbcfproEGFP transfected -3

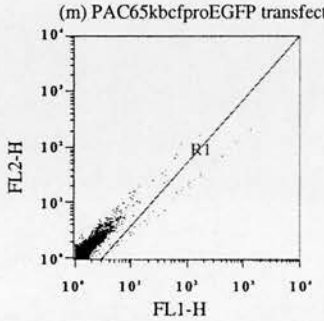


Statistics for: M1KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.72
	(X)				30.18
R1	(Y)	746	7.46	7.46	72.56
	(X)				386.90

C10 Figure 5.2 dotplots (continued)

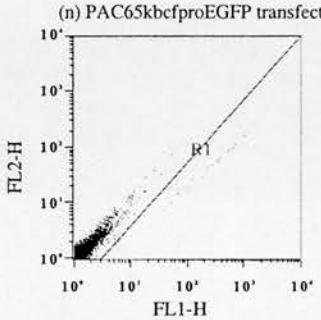
(m) PAC65kbcfproEGFP transfected -1



Statistics for: M65KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.66
	(X)				2.52
R1	(Y)	48	0.48	0.48	39.00
	(X)				216.25

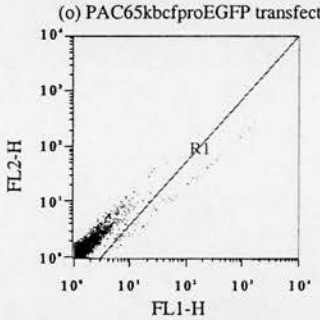
(n) PAC65kbcfproEGFP transfected -2



Statistics for: M65KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.58
	(X)				2.47
R1	(Y)	42	0.42	0.42	44.99
	(X)				257.04

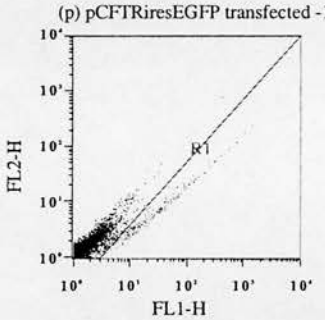
(o) PAC65kbcfproEGFP transfected -3



Statistics for: M65KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.58
	(X)				2.30
R1	(Y)	60	0.60	0.60	27.48
	(X)				150.90

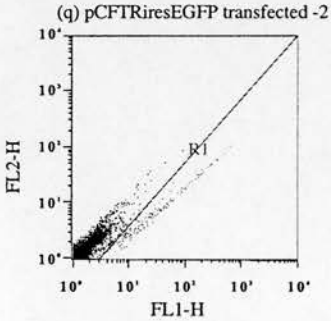
(p) pCFTRiresEGFP transfected -1



Statistics for: MI2ES11 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.98
	(X)				3.59
R1	(Y)	244	2.44	2.44	15.43
	(X)				81.19

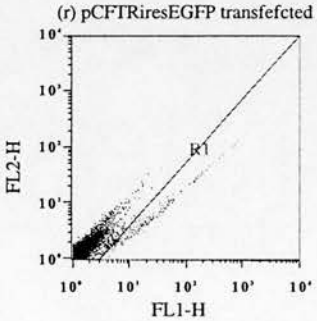
(q) pCFTRiresEGFP transfected -2



Statistics for: MI2ES12 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.73
	(X)				2.68
R1	(Y)	179	1.79	1.79	12.44
	(X)				62.98

(r) pCFTRiresEGFP transfefcted -3

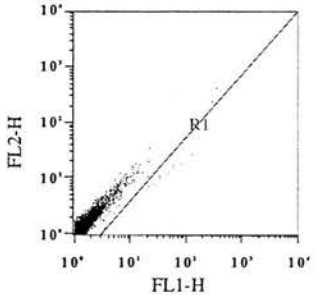


Statistics for: MI2ES13 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.72
	(X)				2.83
R1	(Y)	229	2.29	2.29	11.56
	(X)				59.86

C10 Figure 5.2 dotplots (continued)

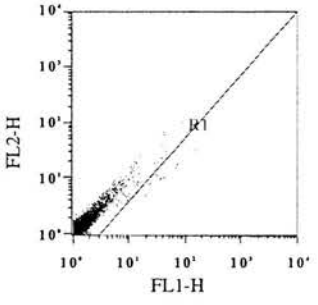
(s) PACRC1iresEGFP transfected -1



Statistics for: MRC1IG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.48
	(X)				1.49
R1	(Y)	15	0.15	0.15	13.14
	(X)				50.78

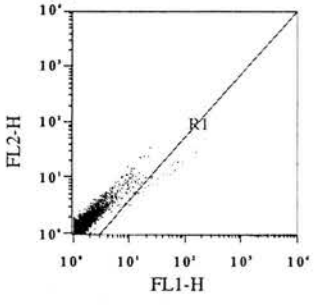
(t) PACRC1iresEGFP transfected -2



Statistics for: MRC1IG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.48
	(X)				1.50
R1	(Y)	20	0.20	0.20	12.71
	(X)				46.69

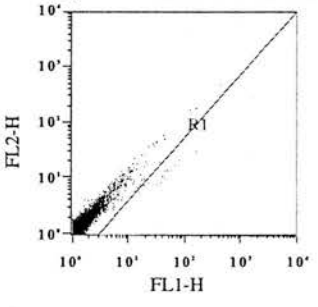
(u) PACRC1iresEGFP transfected -3



Statistics for: MRC1IG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.51
	(X)				1.60
R1	(Y)	29	0.29	0.29	13.66
	(X)				52.99

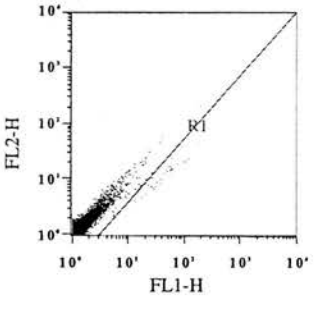
(v) PACRC2iresEGFP transfected -1



Statistics for: MRC2IG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.68
	(X)				1.66
R1	(Y)	26	0.26	0.26	11.37
	(X)				42.82

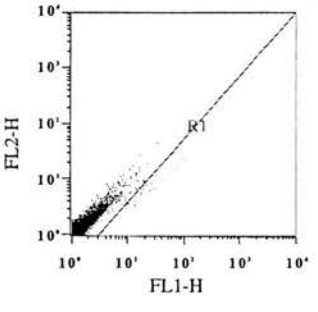
(w) PACRC2iresEGFP transfected -2



Statistics for: MRC2IG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.62
	(X)				1.71
R1	(Y)	32	0.32	0.32	12.91
	(X)				48.69

(x) PACRC2iresEGFP transfected -3

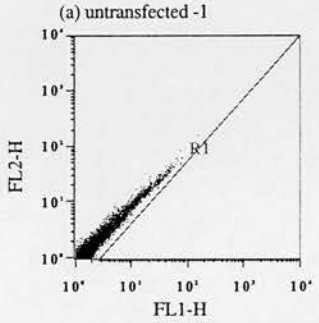


Statistics for: MRC2IG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	1.61
	(X)				1.69
R1	(Y)	33	0.33	0.33	10.18
	(X)				37.02

C11 Figure 5.3 dotplots

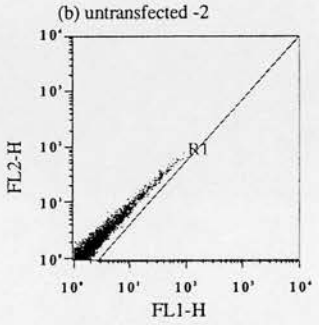
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.73
	(X)				3.09
R1	(Y)	0	0.00	0.00	--
	(X)				--

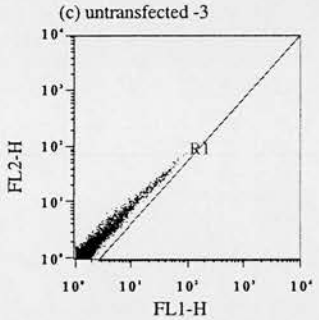
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.32
	(X)				2.59
R1	(Y)	0	0.00	0.00	--
	(X)				--

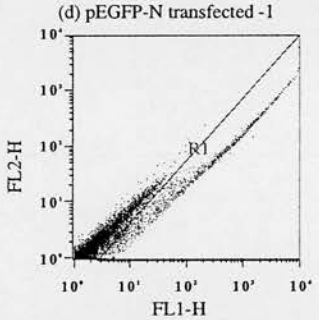
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.23
	(X)				2.51
R1	(Y)	0	0.00	0.00	--
	(X)				--

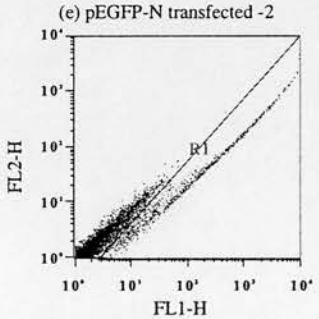
(d) pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	17.01
	(X)				76.64
R1	(Y)	1268	12.68	12.68	116.70
	(X)				582.39

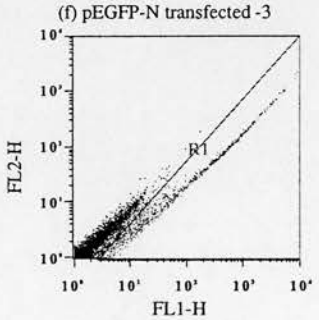
(e) pEGFP-N transfected -2



Statistics for: EGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	13.67
	(X)				59.05
R1	(Y)	1131	11.31	11.31	102.84
	(X)				499.29

(f) pEGFP-N transfected -3

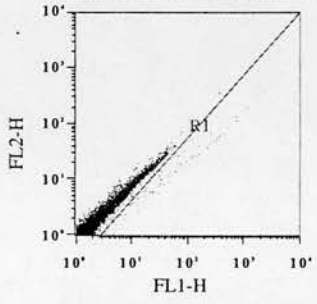


Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.53
	(X)				39.73
R1	(Y)	984	9.84	9.84	78.70
	(X)				381.58

C11 Figure 5.3 dotplots (continued)

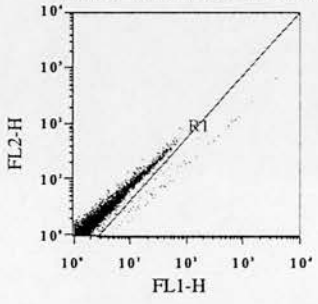
(g) PACRC2cmvEGFP transfected -1



Statistics for: RC2CG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.23
	(X)				6.13
R1	(Y)	112	1.12	1.12	51.31
	(X)				259.60

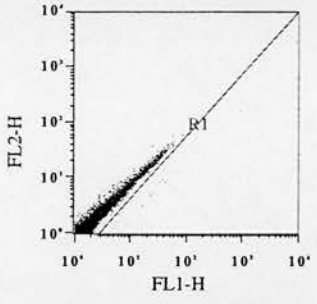
(h) PACRC2cmvEGFP transfected -2



Statistics for: RC2CG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.21
	(X)				8.76
R1	(Y)	88	0.88	0.88	294.29
	(X)				522.23

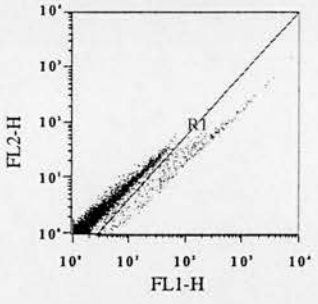
(i) PACRC2cmvEGFP transfected -3



Statistics for: RC2CG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.82
	(X)				4.16
R1	(Y)	20	0.20	0.20	165.08
	(X)				552.83

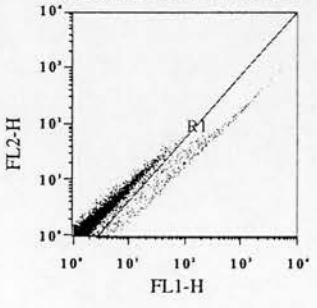
(j) p1kbcfproEGFP transfected -1



Statistics for: 1KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.85
	(X)				14.46
R1	(Y)	465	4.65	4.65	49.50
	(X)				243.59

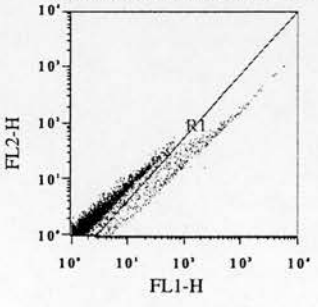
(k) p1kbcfproEGFP transfected -2



Statistics for: 1KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.43
	(X)				13.65
R1	(Y)	533	5.33	5.33	40.19
	(X)				203.42

(l) p1kbcfproEGFP transfected -3

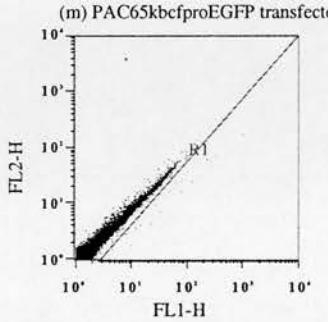


Statistics for: 1KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.98
	(X)				15.94
R1	(Y)	601	6.01	6.01	43.85
	(X)				217.72

C11 Figure 5.3 dotplots (continued)

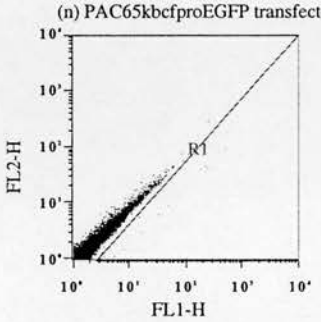
(m) PAC65kbcfproEGFP transfected -1



Statistics for: 65KB4 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	2.69
	(X)				3.34
R1	(Y)	35	0.17	0.17	29.50
	(X)				131.04

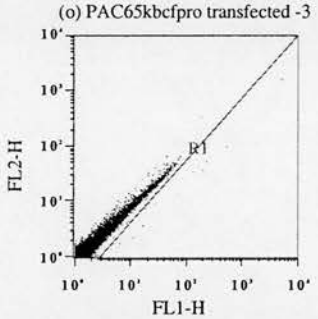
(n) PAC65kbcfproEGFP transfected -2



Statistics for: 65KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	2.00
	(X)				2.37
R1	(Y)	14	0.07	0.07	77.68
	(X)				184.71

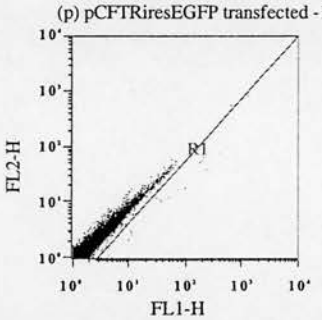
(o) PAC65kbcfpro transfected -3



Statistics for: 65KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	2.36
	(X)				2.85
R1	(Y)	21	0.10	0.10	45.80
	(X)				187.60

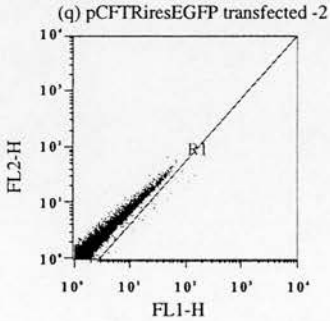
(p) pCFTRiresEGFP transfected -1



Statistics for: I2ES14 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	2.14
	(X)				2.57
R1	(Y)	23	0.12	0.12	56.97
	(X)				132.08

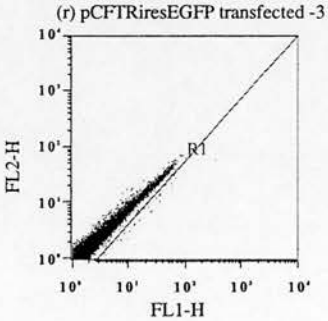
(q) pCFTRiresEGFP transfected -2



Statistics for: I2ES12 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	2.37
	(X)				2.78
R1	(Y)	33	0.17	0.17	16.91
	(X)				57.24

(r) pCFTRiresEGFP transfected -3

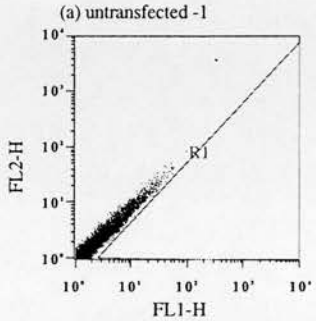


Statistics for: I2ES13 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	20000	100.00	100.00	2.87
	(X)				3.49
R1	(Y)	29	0.14	0.14	11.95
	(X)				37.57

C12 Figure 5.4 dotplots

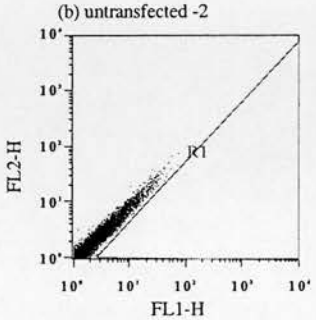
(a) untransfected -1



Statistics for: CND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.72
	(X)				3.21
R1	(Y)	0	0.00	0.00	--
	(X)				--

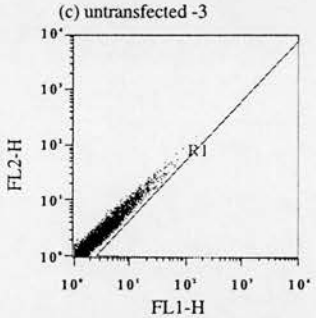
(b) untransfected -2



Statistics for: CND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.66
	(X)				2.96
R1	(Y)	0	0.00	0.00	--
	(X)				--

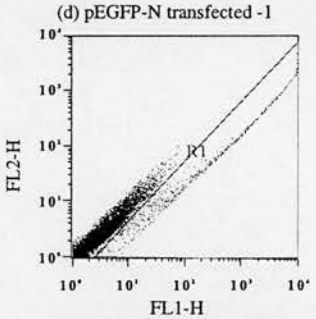
(c) untransfected -3



Statistics for: CND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.70
	(X)				3.01
R1	(Y)	0	0.00	0.00	--
	(X)				--

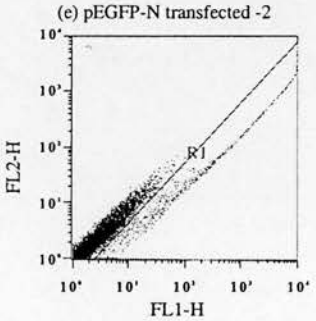
(d) pEGFP-N transfected -1



Statistics for: CEGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	14.66
	(X)				57.50
R1	(Y)	577	5.77	5.77	203.04
	(X)				937.70

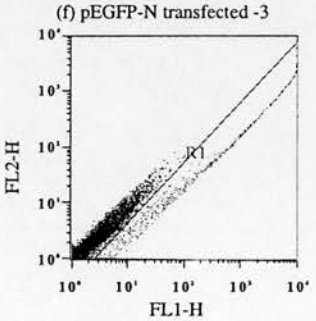
(e) pEGFP-N transfected -2



Statistics for: CEGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	14.58
	(X)				59.40
R1	(Y)	657	6.57	6.57	181.77
	(X)				858.26

(f) pEGFP-N transfected -3

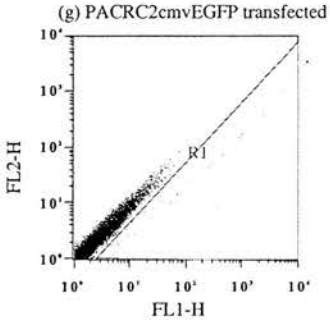


Statistics for: CEGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	16.50
	(X)				64.51
R1	(Y)	582	5.82	5.82	222.36
	(X)				1038.26

C12 Figure 5.4 dotplots (continued)

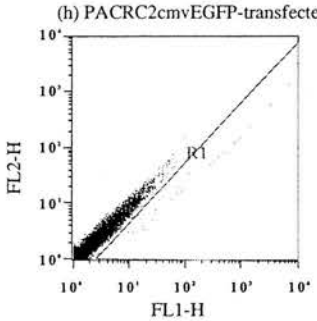
(g) PACRC2cmvEGFP transfected -1



Statistics for: CRC2CG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.25
	(X)				4.82
R1	(Y)	48	0.48	0.48	63.99
	(X)				343.08

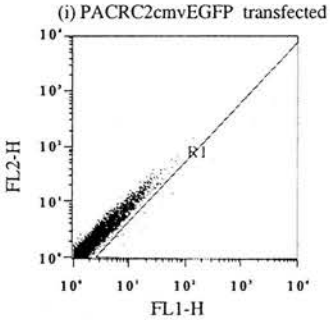
(h) PACRC2cmvEGFP-transfected -2



Statistics for: CRC2CG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.80
	(X)				6.53
R1	(Y)	45	0.45	0.45	172.15
	(X)				726.65

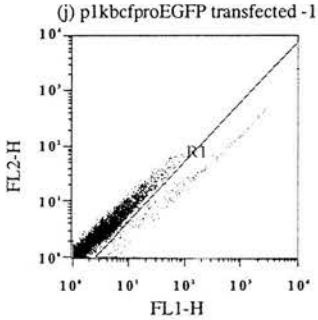
(i) PACRC2cmvEGFP transfected -3



Statistics for: CRC2CG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.87
	(X)				3.81
R1	(Y)	24	0.24	0.24	67.33
	(X)				368.29

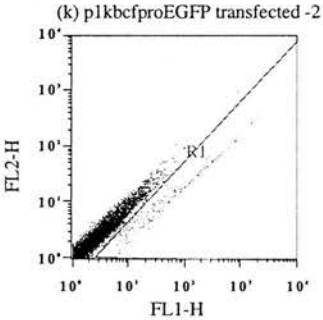
(j) p1kbcfproEGFP transfected -1



Statistics for: C1KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.26
	(X)				9.84
R1	(Y)	205	2.05	2.05	59.86
	(X)				312.39

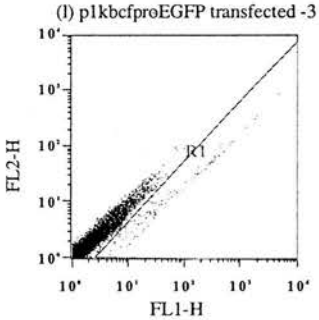
(k) p1kbcfproEGFP transfected -2



Statistics for: C1KB2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.79
	(X)				7.56
R1	(Y)	198	1.98	1.98	40.68
	(X)				217.34

(l) p1kbcfproEGFP transfected -3

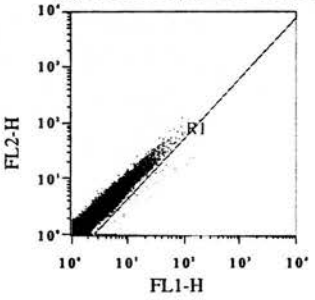


Statistics for: C1KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.17
	(X)				5.93
R1	(Y)	180	1.80	1.80	32.34
	(X)				173.63

C12 Figure 5.4 dotplots (continued)

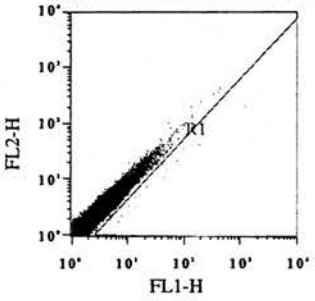
(m) PAC65kbcfproEGFP transfected -1



Statistics for: C65KB1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	50000	100.00	100.00	2.74
	(X)				2.99
R1	(Y)	43	0.09	0.09	9.21
	(X)				36.10

(n) PAC65kbcfproEGFP transfected -2

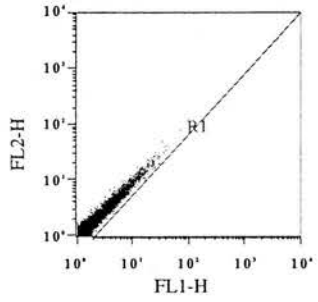


Statistics for: C65KB3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	50000	100.00	100.00	2.84
	(X)				3.11
R1	(Y)	25	0.05	0.05	21.33
	(X)				103.30

C13 Figure 5.5 dotplots

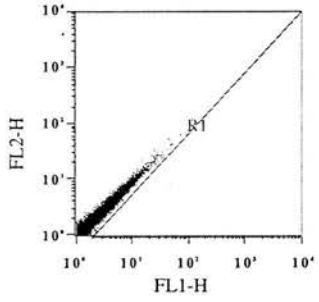
(a) untransfected -1



Statistics for: HND1A , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.84
	(X)				3.02
R1	(Y)	0	0.00	0.00	--
	(X)				--

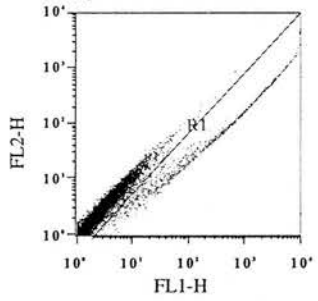
(b) untransfected -2



Statistics for: HND1B , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.21
	(X)				3.43
R1	(Y)	0	0.00	0.00	--
	(X)				--

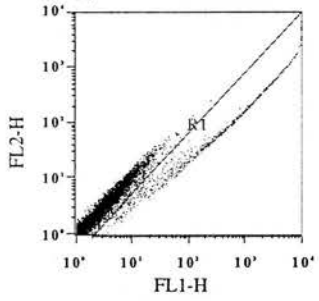
(c) pEGFP-N transfected -1



Statistics for: HEGFPNA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	24.01
	(X)				102.74
R1	(Y)	1160	11.60	11.60	174.39
	(X)				854.76

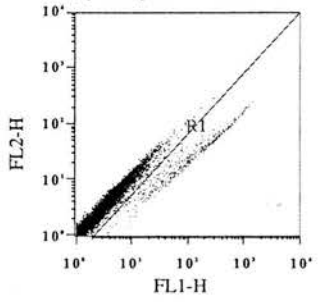
(d) pEGFP-N transfected -2



Statistics for: HEGFPNB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	25.22
	(X)				105.73
R1	(Y)	1040	10.40	10.40	207.38
	(X)				983.11

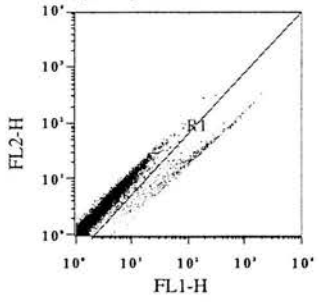
(e) p1kbcfproEGFP transfected -1



Statistics for: H1KBA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.56
	(X)				14.52
R1	(Y)	604	6.04	6.04	34.21
	(X)				172.11

(f) p1kbcpfroEGFP transfected -2

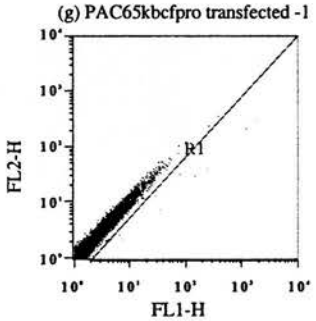


Statistics for: H1KBB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.63
	(X)				11.79
R1	(Y)	501	5.01	5.01	31.49
	(X)				161.84

C13 Figure 5.5 dotplots (continued)

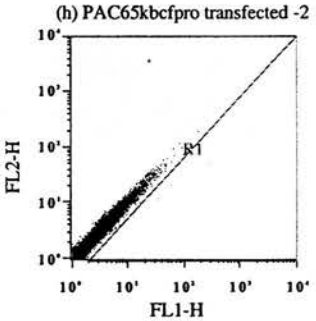
(g) PAC65kbcfpro transfected -1



Statistics for: H65KBA , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.90
	(X)				3.92
R1	(Y)	11	0.11	0.11	62.43
	(X)				328.90

(h) PAC65kbcfpro transfected -2

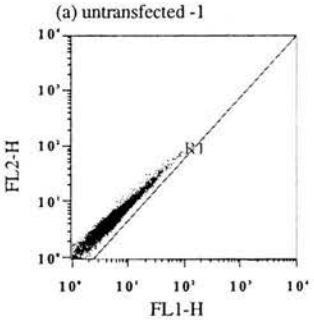


Statistics for: H65KBB , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.91
	(X)				3.68
R1	(Y)	4	0.04	0.04	26.20
	(X)				117.97

C14 Figure 5.6 dotplots

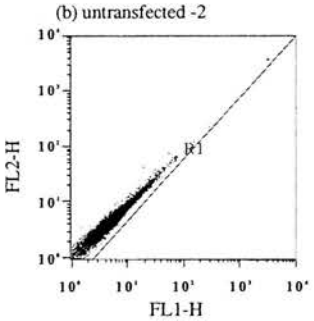
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.75
	(X)				6.08
R1	(Y)	0	0.00	0.00	--
	(X)				--

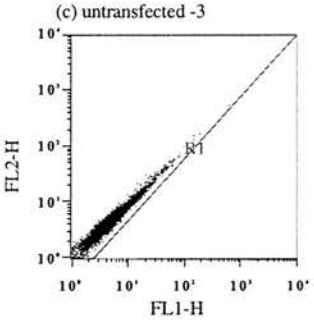
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.75
	(X)				6.12
R1	(Y)	0	0.00	0.00	--
	(X)				--

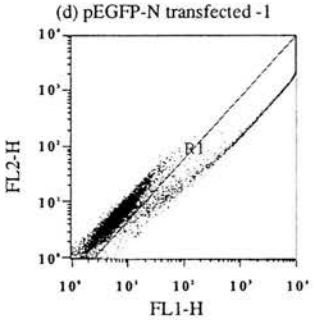
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.66
	(X)				5.98
R1	(Y)	0	0.00	0.00	--
	(X)				--

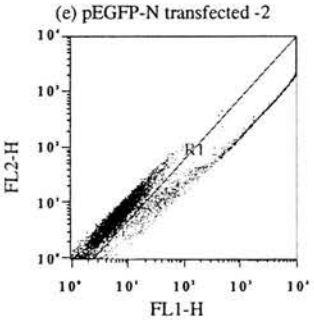
(d) pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	297.45
	(X)				1155.22
R1	(Y)	2478	24.78	24.78	1160.94
	(X)				4621.31

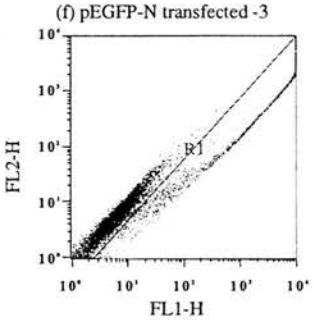
(e) pEGFP-N transfected -2



Statistics for: EGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	252.56
	(X)				983.71
R1	(Y)	2218	22.18	22.18	1083.01
	(X)				4377.82

(f) pEGFP-N transfected -3

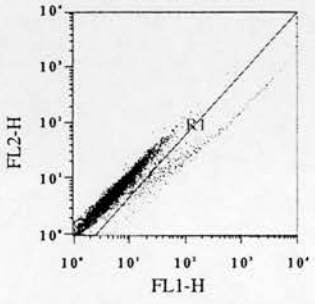


Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	273.55
	(X)				1053.28
R1	(Y)	2189	21.89	21.89	1198.77
	(X)				4760.16

C14 Figure 5.6 dotplots (continued)

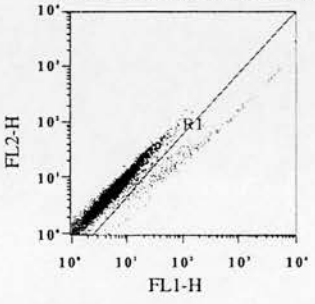
(g) PACRC2cmvEGFP transfected -1



Statistics for: RC2CGFP1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	11.78
	(X)				24.25
R1	(Y)	393	3.93	3.93	121.62
	(X)				454.09

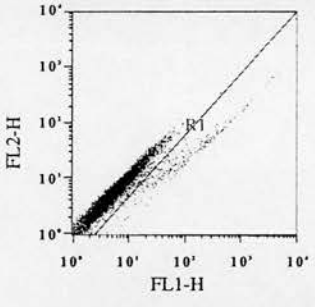
(h) PACRC2cmvEGFP transfected -2



Statistics for: RC2CGFP2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	12.04
	(X)				26.49
R1	(Y)	385	3.85	3.85	136.67
	(X)				526.78

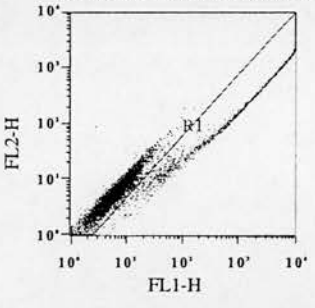
(i) PACRC2cmvEGFP transfected -3



Statistics for: RC2CGFP3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.66
	(X)				17.04
R1	(Y)	316	3.16	3.16	96.04
	(X)				343.59

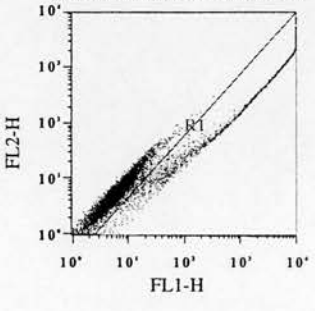
(j) pEGFP-N/PUC18 transfected -1



Statistics for: -NPUC181 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	451.95
	(X)				1809.05
R1	(Y)	3494	34.94	34.94	1258.21
	(X)				5141.99

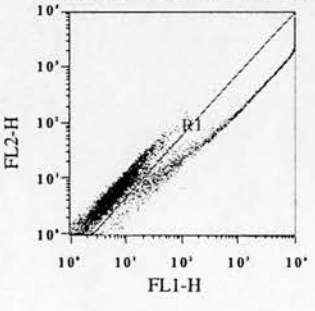
(k) pEGFP-N/PUC18 tranfected -2



Statistics for: -NPUC182 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	379.33
	(X)				1513.84
R1	(Y)	3137	31.37	31.37	1179.07
	(X)				4795.25

(l) pEGFP-N/PUC18 transfected -3

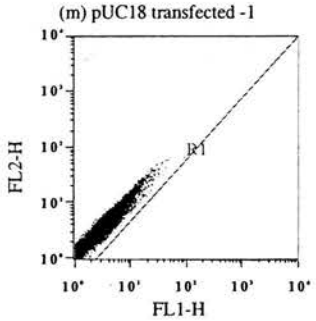


Statistics for: -NPUC183 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	385.06
	(X)				1520.86
R1	(Y)	3151	31.51	31.51	1178.54
	(X)				4782.89

C14 Figure 5.6 dotplots (continued)

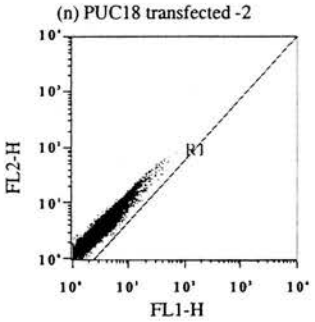
(m) pUC18 transfected -1



Statistics for: 1UGPUC181 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.87
	(X)				3.91
R1	(Y)	0	0.00	0.00	--
	(X)				--

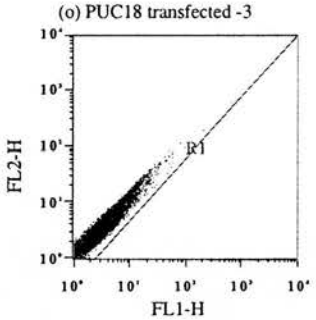
(n) PUC18 transfected -2



Statistics for: 1UGPUC182 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.97
	(X)				3.99
R1	(Y)	0	0.00	0.00	--
	(X)				--

(o) PUC18 transfected -3

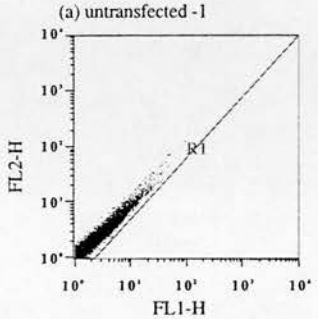


Statistics for: 1UGPUC183 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.90
	(X)				3.99
R1	(Y)	0	0.00	0.00	--
	(X)				--

C15 Figure 5.7 dotplots

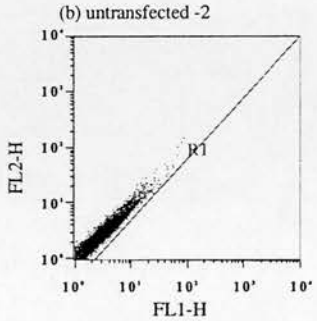
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.15
	(X)				3.34
R1	(Y)	0	0.00	0.00	--
	(X)				--

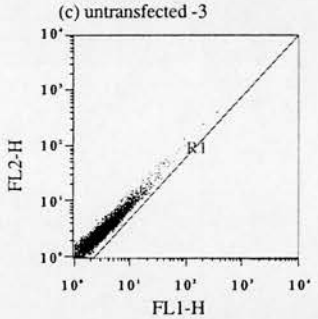
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.07
	(X)				3.22
R1	(Y)	0	0.00	0.00	--
	(X)				--

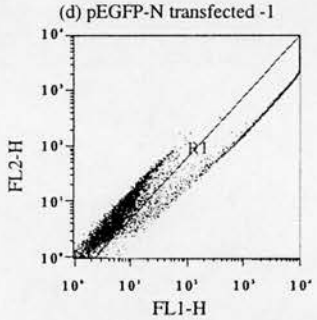
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.21
	(X)				3.35
R1	(Y)	0	0.00	0.00	--
	(X)				--

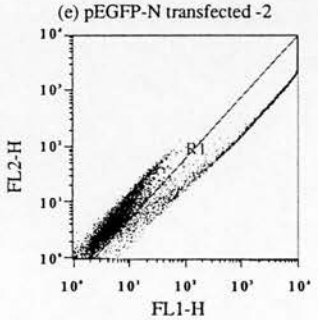
(d) pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	433.42
	(X)				1541.44
R1	(Y)	3209	32.09	32.09	1282.03
	(X)				4736.46

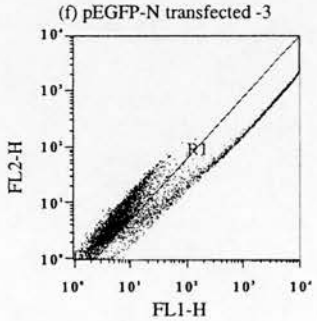
(e) pEGFP-N transfected -2



Statistics for: EGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	445.23
	(X)				1600.53
R1	(Y)	3635	36.35	36.35	1169.49
	(X)				4348.25

(f) pEGFP-N transfected -3

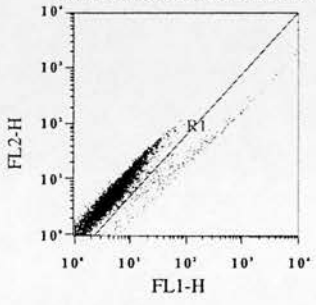


Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	536.02
	(X)				1889.66
R1	(Y)	4384	43.84	43.84	1144.62
	(X)				4232.38

C15 Figure 5.7 dotplots (continued)

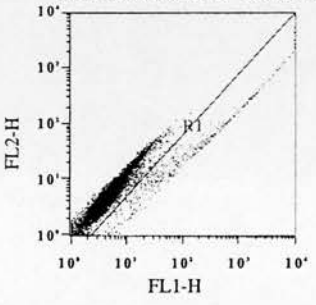
(g) PACRC2cmvEGFP transfected -1



Statistics for: RC2CG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	15.88
	(X)				29.37
R1	(Y)	357	3.57	3.57	165.90
	(X)				570.62

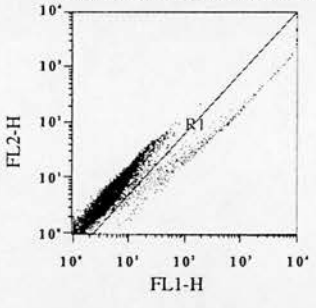
(h) PACRC2cmvEGFP transfected -2



Statistics for: RC2CG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	24.17
	(X)				56.45
R1	(Y)	586	5.86	5.86	255.15
	(X)				824.97

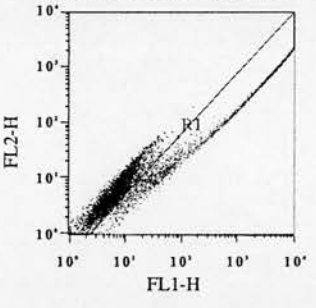
(i) PACRC2cmvEGFP transfected -3



Statistics for: RC2CG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	24.15
	(X)				54.06
R1	(Y)	511	5.11	5.11	257.07
	(X)				860.97

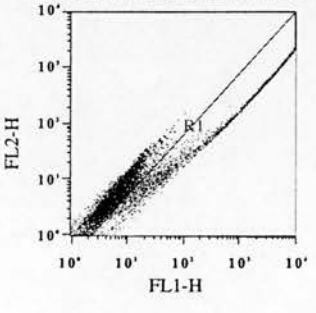
(j) pEGFP-N/RC2b transfected -1



Statistics for: -N+RC2B1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	483.14
	(X)				1684.85
R1	(Y)	3684	36.84	36.84	1257.28
	(X)				4518.52

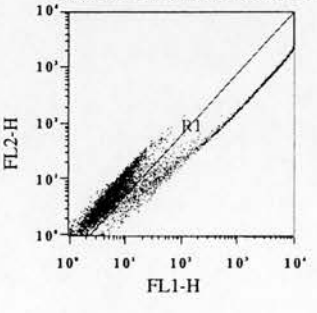
(k) pEGFP-N/RC2 transfected -2



Statistics for: -N+RC2B2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	500.96
	(X)				1794.98
R1	(Y)	4100	41.00	41.00	1171.68
	(X)				4327.34

(l) pEGFP-N/RC2b transfected -3

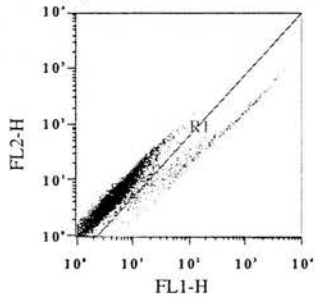


Statistics for: -N+RC2B3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	562.99
	(X)				1966.15
R1	(Y)	4063	40.63	40.63	1323.81
	(X)				4777.08

C15 Figure 5.7 dotplots (continued)

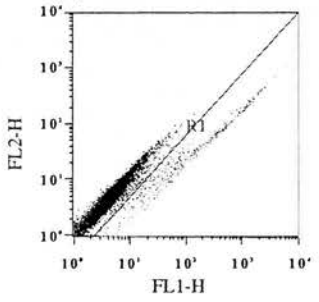
(m) PACRC2cmvEGFP/RC2B transfected -1



Statistics for: CG+RC2B1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.89
	(X)				22.31
R1	(Y)	551	5.51	5.51	63.23
	(X)				321.42

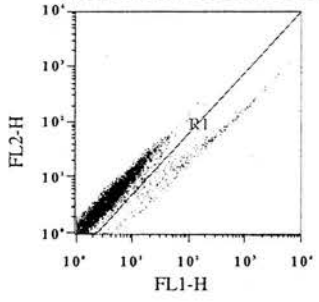
(n) PACRC2cmvEGFP/RC2B transfected -2



Statistics for: CG+RC2B2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.52
	(X)				21.07
R1	(Y)	534	5.34	5.34	61.09
	(X)				310.71

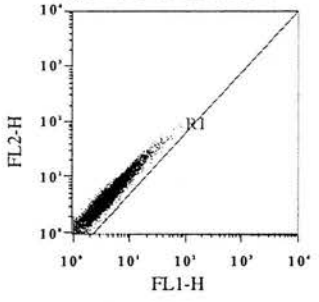
(o) PACRC2cmvEGFP/RC2B transfected -3



Statistics for: CG+RC2B3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.00
	(X)				14.97
R1	(Y)	389	3.89	3.89	53.96
	(X)				275.75

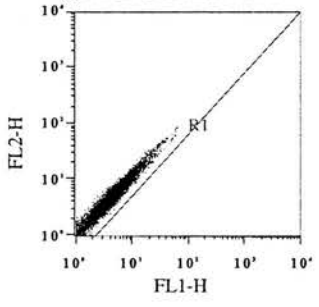
(p) PACRC2b transfected -1



Statistics for: RC2B1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.23
	(X)				4.47
R1	(Y)	0	0.00	0.00	--
	(X)				--

(q) PACRC2b transfected -2

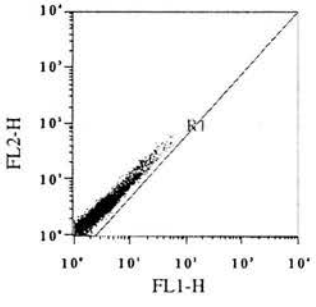


Statistics for: RC2B2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.38
	(X)				4.55
R1	(Y)	0	0.00	0.00	--
	(X)				--

C16 Figure 5.8 dotplots

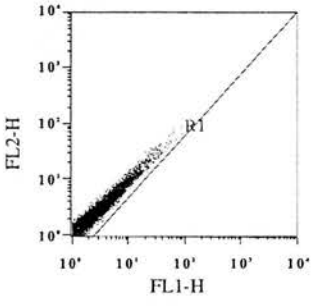
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.40
	(X)				3.51
R1	(Y)	0	0.00	0.00	--
	(X)				--

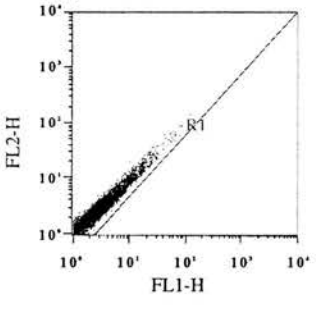
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.38
	(X)				3.44
R1	(Y)	0	0.00	0.00	--
	(X)				--

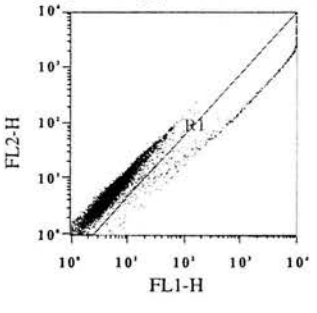
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	3.42
	(X)				3.50
R1	(Y)	0	0.00	0.00	--
	(X)				--

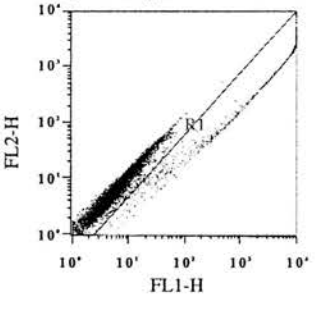
(d) 1 microg pEGFP-N transfected -1



Statistics for: 1UG-N1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	101.41
	(X)				342.30
R1	(Y)	881	8.81	8.81	1034.97
	(X)				3780.40

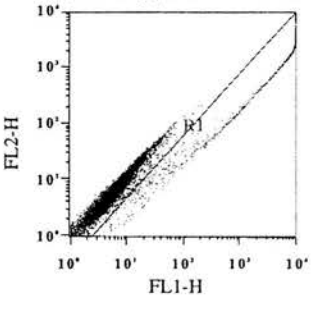
(e) 1 microg pEGFP-N transfected -2



Statistics for: 1UG-N2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	108.44
	(X)				360.70
R1	(Y)	906	9.06	9.06	1084.06
	(X)				3878.80

(f) 1 microg pEGFP-N transfected -3

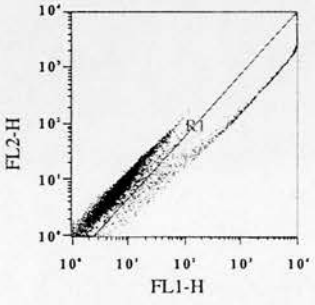


Statistics for: 1UG-N3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	102.31
	(X)				329.88
R1	(Y)	871	8.71	8.71	1042.78
	(X)				3663.08

C16 Figure 5.8 dotplots (continued)

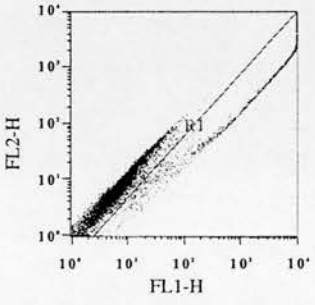
(g) 100ng pEGFP-N + 900 ng pUC18 transfected -1



Statistics for: 100NG-N1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	197.26
	(X)				703.80
R1	(Y)	1702	17.02	17.02	1115.35
	(X)				4094.96

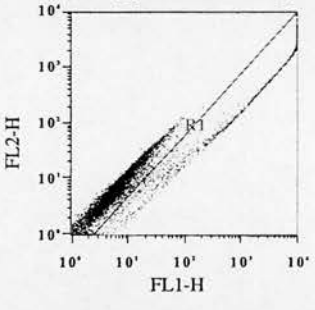
(h) 100ng pEGFP-N + 900 ng pUC18 transfected -2



Statistics for: 100NG-N2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	185.66
	(X)				632.55
R1	(Y)	1575	15.75	15.75	1091.50
	(X)				3933.16

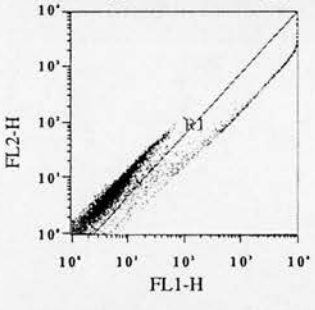
(i) 100ng pEGFP-N + 900 ng pUC18 transfected -3



Statistics for: 100NG-N3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	139.33
	(X)				505.98
R1	(Y)	1402	14.02	14.02	925.39
	(X)				3546.41

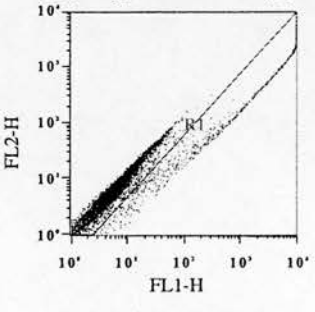
(j) 10ng pEGFP-N + 990 ng pUC18 transfected -1



Statistics for: 10NG-N1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	72.47
	(X)				279.27
R1	(Y)	1117	11.17	11.17	591.14
	(X)				2447.67

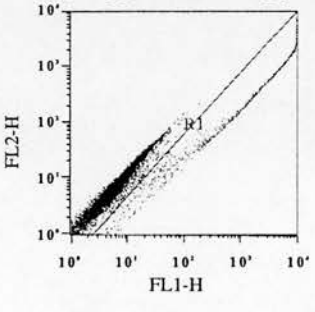
(k) 10ng pEGFP-N + 990 ng pUC18 transfected -2



Statistics for: 10NG-N2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	117.62
	(X)				416.12
R1	(Y)	1105	11.05	11.05	974.91
	(X)				3684.45

(l) 10ng pEGFP-N + 990 ng pUC18 transfected -3

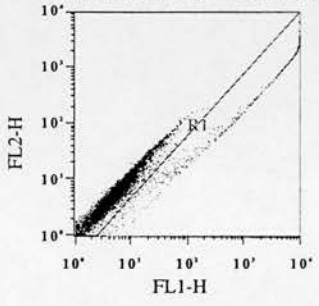


Statistics for: 10NG-N3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	78.70
	(X)				281.68
R1	(Y)	928	9.28	9.28	743.09
	(X)				2942.64

C16 Figure 5.8 dotplots (continued)

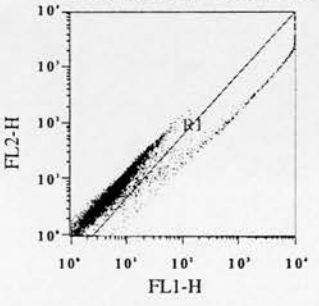
(m) 1ng pEGFP-N + 999 ng pUC18 transfected -1



Statistics for: 1NG-N1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	53.34
	(X)				198.29
R1	(Y)	665	6.65	6.65	710.23
	(X)				2901.21

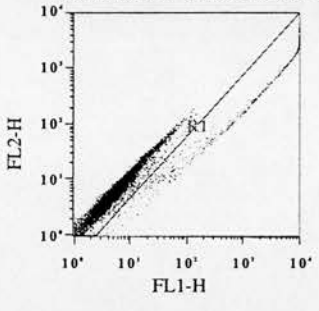
(n) 1ng pEGFP-N + 999 ng pUC18 transfected -2



Statistics for: 1NG-N2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	75.34
	(X)				254.85
R1	(Y)	736	7.36	7.36	908.95
	(X)				3358.67

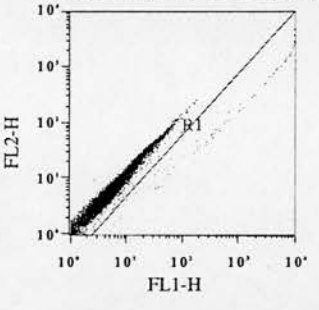
(o) 1ng pEGFP-N + 999 ng pUC18 transfected -3



Statistics for: 1NG-N3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	73.60
	(X)				230.98
R1	(Y)	627	6.27	6.27	1006.95
	(X)				3537.26

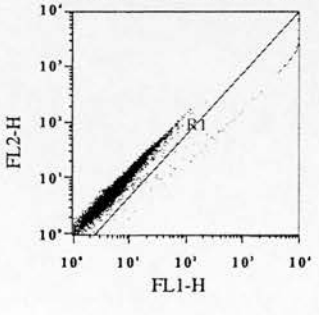
(p) 100pg pEGFP-N + 999.9 ng pUC18 transfected -1



Statistics for: 100PG-N2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	16.95
	(X)				34.96
R1	(Y)	111	1.11	1.11	693.02
	(X)				2490.42

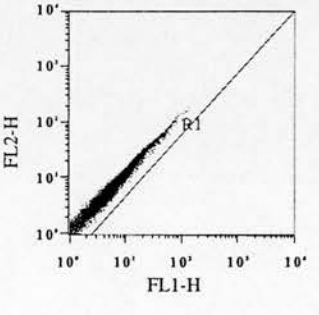
(q) 100pg pEGFP-N + 999.9 ng pUC18 transfected -2



Statistics for: 100PG-N3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	19.07
	(X)				40.86
R1	(Y)	104	1.04	1.04	938.62
	(X)				3192.61

(r) 1 microg PUC18 transfected -1

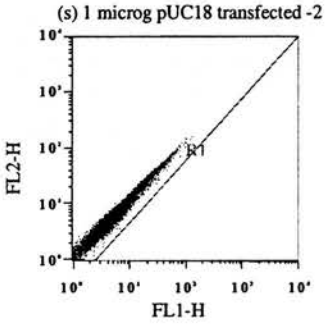


Statistics for: PUC181 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.91
	(X)				6.87
R1	(Y)	0	0.00	0.00	--
	(X)				--

C16 Figure 5.8 dotplots (continued)

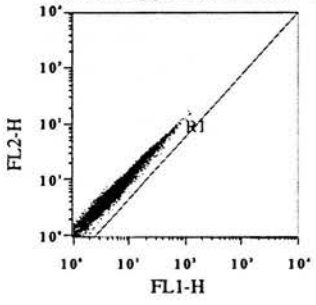
(s) 1 microg pUC18 transfected -2



Statistics for: PUC182 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.88
	(X)				6.87
R1	(Y)	0	0.00	0.00	--
	(X)				--

(t) 1 microg pUC18 transfected -3

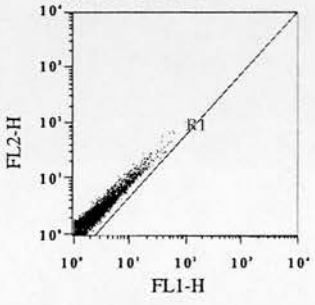


Statistics for: PUC183 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.07
	(X)				7.04
R1	(Y)	0	0.00	0.00	--
	(X)				--

C17 Figure 5.9 dotplots

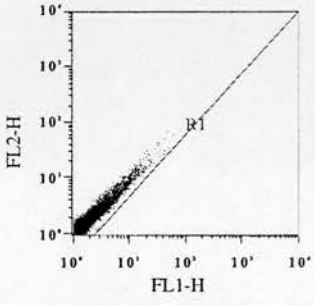
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	2.72
(X)				2.80
R1 (Y)	0	0.00	0.00	--
(X)				--

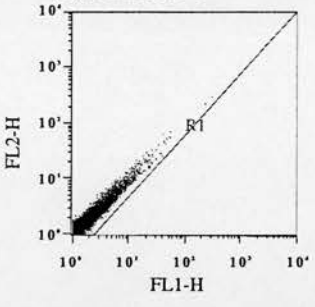
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	2.56
(X)				2.60
R1 (Y)	0	0.00	0.00	--
(X)				--

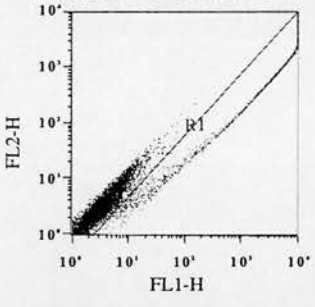
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	2.68
(X)				2.69
R1 (Y)	0	0.00	0.00	--
(X)				--

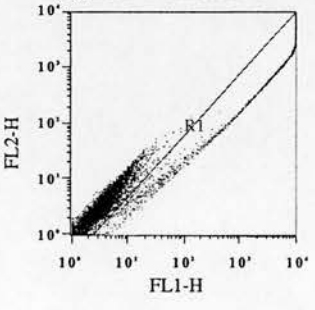
(d) pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	228.68
(X)				812.96
R1 (Y)	1803	18.03	18.03	1201.52
(X)				4442.69

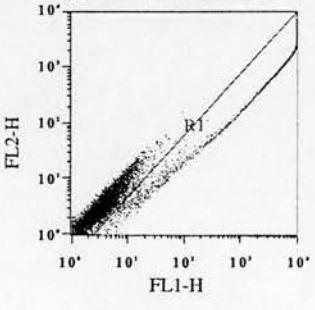
(e) pEGFP-N transfected -2



Statistics for: EGFPN2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	277.06
(X)				1002.38
R1 (Y)	2184	21.84	21.84	1213.53
(X)				4535.17

(f) pEGFP-N transfected -3

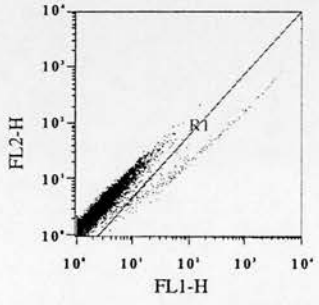


Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	244.61
(X)				881.68
R1 (Y)	2025	20.25	20.25	1173.71
(X)				4319.94

C17 Figure 5.9 dotplots (continued)

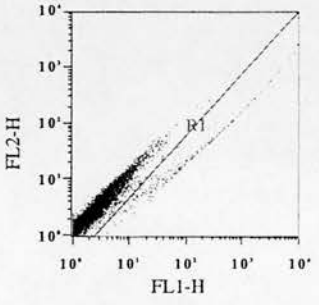
(g) PACRC2cmvEGFP transfected -1



Statistics for: RC2CG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	9.56
	(X)				20.92
R1	(Y)	244	2.44	2.44	229.76
	(X)				727.64

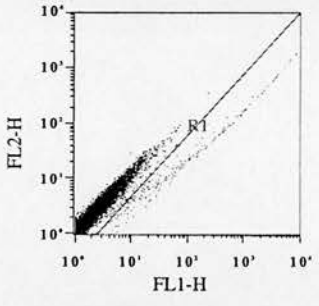
(h) PACRC2cmvEGFP transfected -2



Statistics for: RC2CG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	10.53
	(X)				26.25
R1	(Y)	346	3.46	3.46	191.93
	(X)				668.09

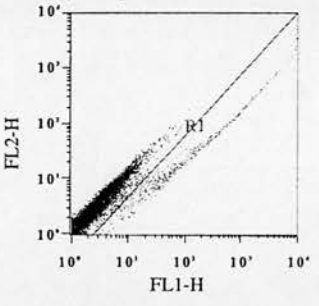
(i) PACRC2cmvEGFP transfected -3



Statistics for: RC2CG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.46
	(X)				18.45
R1	(Y)	268	2.68	2.68	134.42
	(X)				532.38

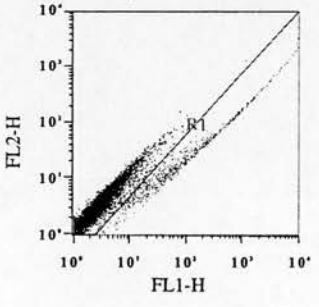
(j) 100ng PACRC2cmvEGFP + 900 ng pUC18 transfected -1



Statistics for: 100NG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	10.91
	(X)				32.34
R1	(Y)	567	5.67	5.67	126.74
	(X)				516.13

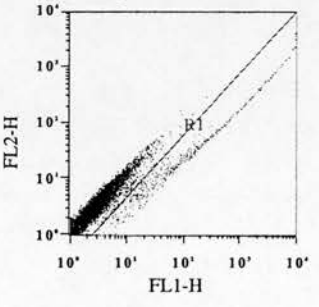
(k) 100ng PACRC2cmvEGFP + 900 ng pUC18 transfected -2



Statistics for: 100NG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	15.93
	(X)				46.89
R1	(Y)	750	7.50	7.50	159.37
	(X)				581.18

(l) 100ng PACRC2cmvEGFP + 900 ng pUC18 transfected -3

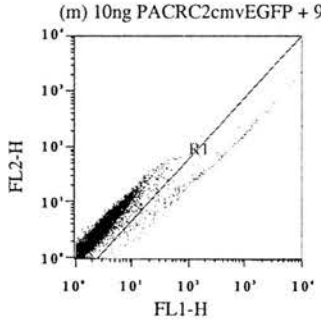


Statistics for: 100NG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	12.64
	(X)				43.17
R1	(Y)	558	5.58	5.58	160.33
	(X)				718.48

C17 Figure 5.9 dotplots (continued)

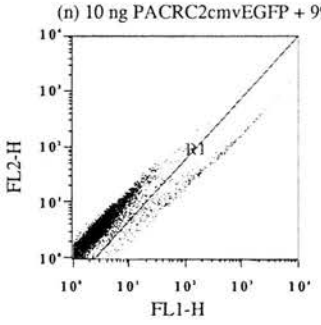
(m) 10ng PACRC2cmvEGFP + 990 ng pUC18 -1



Statistics for: 10NG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.62
	(X)				13.28
R1	(Y)	285	2.85	2.85	71.63
	(X)				360.61

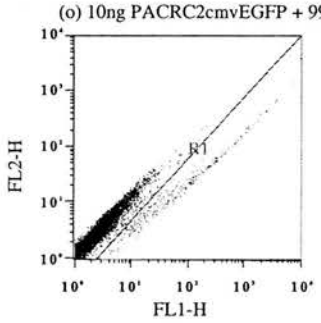
(n) 10 ng PACRC2cmvEGFP + 990ng pUC18 -2



Statistics for: 10NG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	7.29
	(X)				20.31
R1	(Y)	477	4.77	4.77	68.29
	(X)				356.50

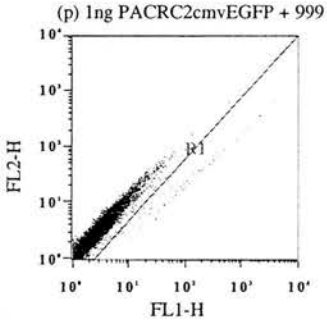
(o) 10ng PACRC2cmvEGFP + 990 ng pUC18 -3



Statistics for: 10NG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.14
	(X)				14.12
R1	(Y)	390	3.90	3.90	61.08
	(X)				282.61

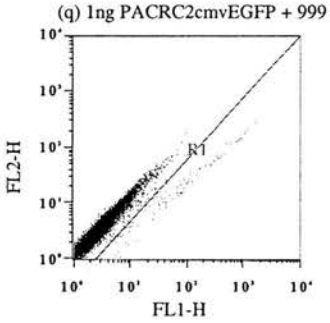
(p) 1ng PACRC2cmvEGFP + 999 ng pUC18 -1



Statistics for: 1NG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.73
	(X)				6.41
R1	(Y)	98	0.98	0.98	57.61
	(X)				301.20

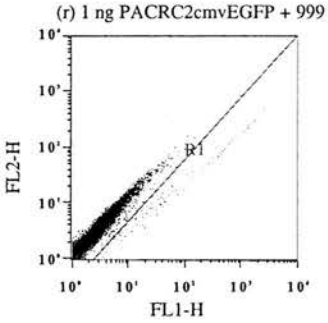
(q) 1ng PACRC2cmvEGFP + 999 ng pUC18 -2



Statistics for: 1NG2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.34
	(X)				7.24
R1	(Y)	144	1.44	1.44	49.03
	(X)				252.49

(r) 1 ng PACRC2cmvEGFP + 999 ng pUC18 -3

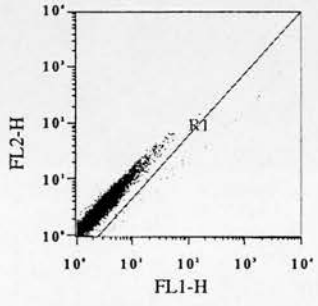


Statistics for: 1NG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.88
	(X)				6.60
R1	(Y)	126	1.26	1.26	49.35
	(X)				251.70

C17 Figure 5.9 dotplots (continued)

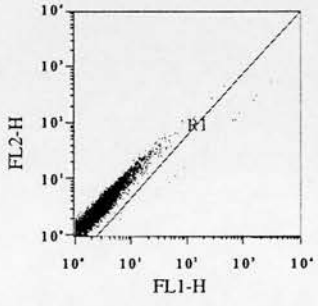
(s) 100pg PACRC2cmvEGFP + 999.9 ng pUC18 -1



Statistics for: 100PG1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.24
	(X)				4.68
R1	(Y)	32	0.32	0.32	80.99
	(X)				427.08

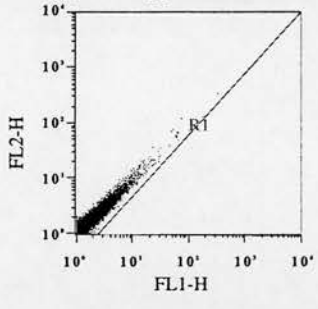
(t) 100pg PACRC2cmvEGFP + 999.9 ng pUC18 -2



Statistics for: 100PG3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.16
	(X)				4.73
R1	(Y)	23	0.23	0.23	118.30
	(X)				647.57

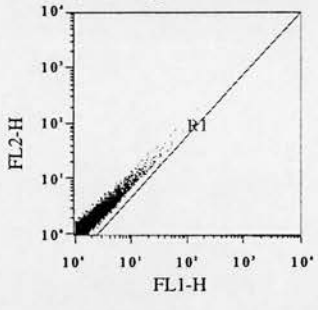
(u) 1 microg pUC18 transfected-1



Statistics for: PUC181 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.57
	(X)				2.56
R1	(Y)	0	0.00	0.00	--
	(X)				--

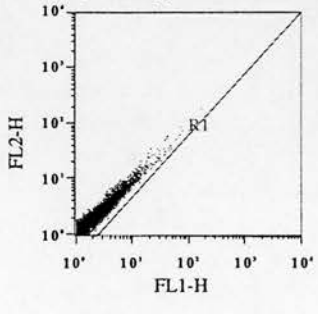
(v) 1 microg pUC18 transfected -2



Statistics for: PUC182 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.59
	(X)				2.72
R1	(Y)	0	0.00	0.00	--
	(X)				--

(w) 1 microg pUC18 transfected -3

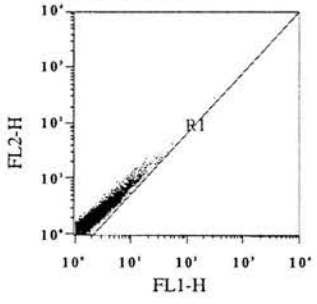


Statistics for: PUC183 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.64
	(X)				2.70
R1	(Y)	0	0.00	0.00	--
	(X)				--

C18 Figure 5.10 dotplots

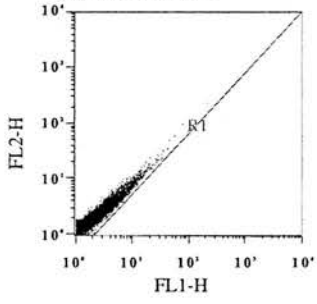
(a) untransfected -1



Statistics for: ND1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.79
	(X)				3.06
R1	(Y)	0	0.00	0.00	--
	(X)				--

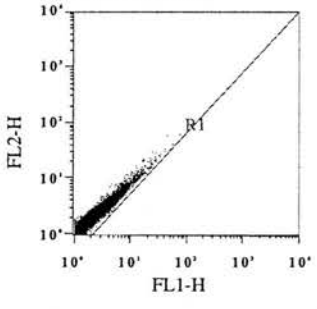
(b) untransfected -2



Statistics for: ND2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.78
	(X)				2.98
R1	(Y)	0	0.00	0.00	--
	(X)				--

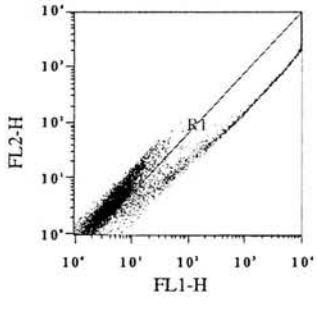
(c) untransfected -3



Statistics for: ND3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	2.85
	(X)				3.10
R1	(Y)	0	0.00	0.00	--
	(X)				--

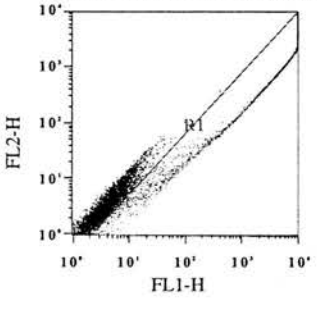
(d) 1 microg pEGFP-N transfected -1



Statistics for: EGFPN1 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	416.58
	(X)				1505.56
R1	(Y)	3257	32.57	32.57	1238.88
	(X)				4581.49

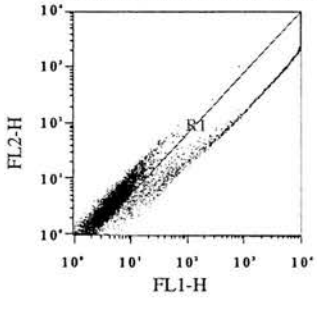
(e) 1 microg pEGFP-N transfected -2



Statistics for: EGFPN2 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	359.79
	(X)				1276.30
R1	(Y)	2703	27.03	27.03	1264.35
	(X)				4654.45

(f) 1 microg pEGFP-N transfected -3

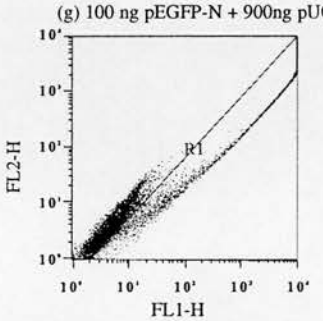


Statistics for: EGFPN3 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	352.31
	(X)				1278.23
R1	(Y)	2879	28.79	28.79	1181.42
	(X)				4396.00

C18 Figure 5.10 dotplots (continued)

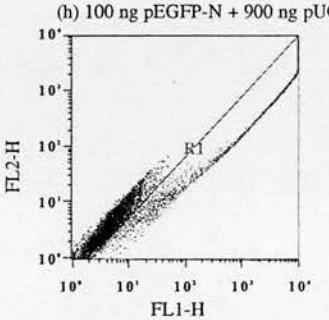
(g) 100 ng pEGFP-N + 900ng pUC18, mixed -1



Statistics for: MIX1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	402.80
(X)				1481.41
R1 (Y)	3420	34.20	34.20	1140.05
(X)				4292.76

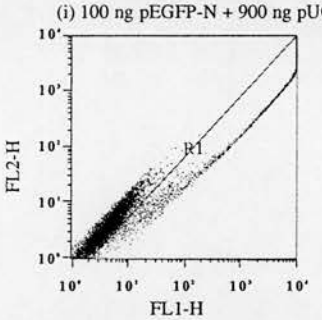
(h) 100 ng pEGFP-N + 900 ng pUC18, mixed -2



Statistics for: MIX2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	459.89
(X)				1674.66
R1 (Y)	3484	34.84	34.84	1248.08
(X)				4734.39

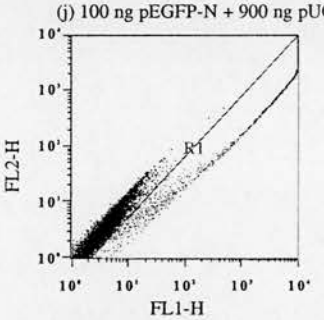
(i) 100 ng pEGFP-N + 900 ng pUC18, mixed -3



Statistics for: MIX3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	399.00
(X)				1459.20
R1 (Y)	2969	29.69	29.69	1282.66
(X)				4853.13

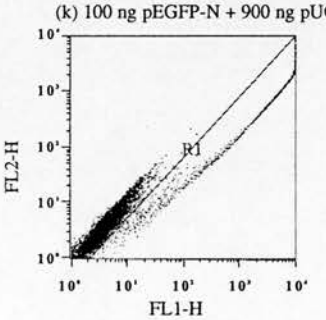
(j) 100 ng pEGFP-N + 900 ng pUC18, separate -1



Statistics for: SEP1 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	161.63
(X)				551.33
R1 (Y)	1415	14.15	14.15	1071.43
(X)				3826.01

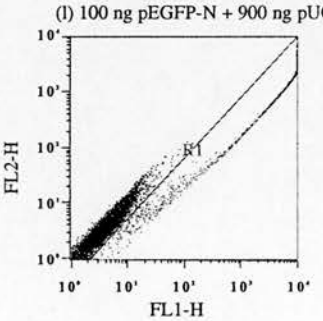
(k) 100 ng pEGFP-N + 900 ng pUC18, separate -2



Statistics for: SEP2 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	148.80
(X)				558.57
R1 (Y)	1434	14.34	14.34	996.39
(X)				3854.94

(l) 100 ng pEGFP-N + 900 ng pUC18, separate -3

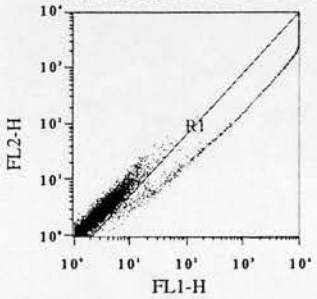


Statistics for: SEP3 , FL2-H v FL1-H , Ungated

Region	Count	% Gated	% Total	Mean
Total (Y)	10000	100.00	100.00	146.28
(X)				537.53
R1 (Y)	1307	13.07	13.07	1052.27
(X)				4045.92

C18 Figure 5.10 dotplots (continued)

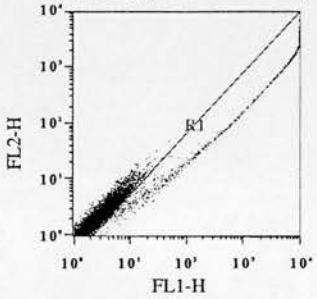
(m) 100 ng pEGFP-N + 0.75 Lipofectin -1



Statistics for: 100.751 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	167.15
	(X)				577.34
R1	(Y)	1242	12.42	12.42	1307.18
	(X)				4607.84

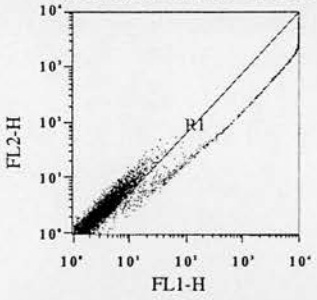
(n) 199 ng pEGFP-N + 0.75 Lipofectin -2



Statistics for: 100.752 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	99.41
	(X)				371.69
R1	(Y)	1136	11.36	11.36	835.75
	(X)				3227.32

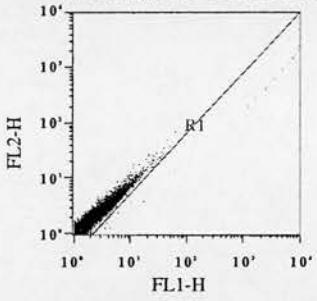
(o) 100 ng pEGFP-N + 0.75 Lipofectin -3



Statistics for: 100.753 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	134.48
	(X)				478.16
R1	(Y)	1234	12.34	12.34	1051.82
	(X)				3832.60

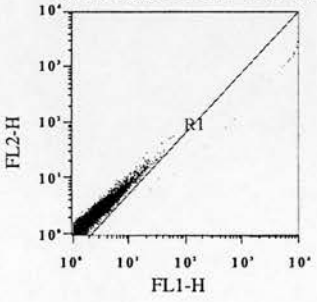
(p) 100 ng pEGFP-N + 0.075 Lipofectin -1



Statistics for: 100.0751 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	6.00
	(X)				13.98
R1	(Y)	28	0.28	0.28	1203.21
	(X)				3958.65

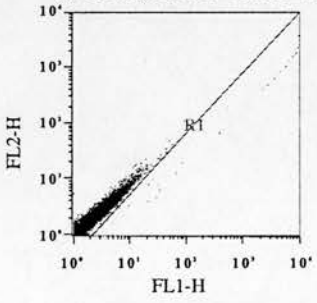
(q) 100 ng pEGFP-N + 0.075 Lipofectin -2



Statistics for: 100.0752 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	8.41
	(X)				24.54
R1	(Y)	38	0.38	0.38	1482.52
	(X)				5656.24

(r) 100 ng pEGFP-N + 0.075 Lipofectin -3

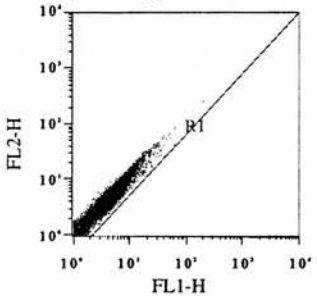


Statistics for: 100.0753 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.07
	(X)				12.77
R1	(Y)	34	0.34	0.34	694.57
	(X)				2889.86

C18 Figure 5.10 dotplots (continued)

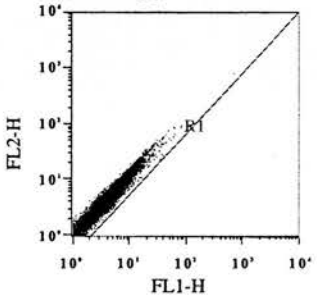
(s) 1 microg pUC18 transfected -1



Statistics for: PUC181 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.53
	(X)				3.90
R1	(Y)	0	0.00	0.00	--
	(X)				--

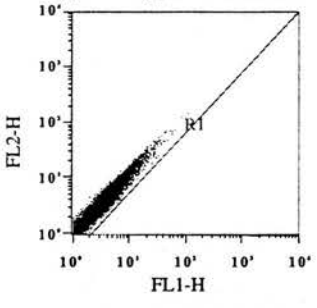
(t) 1 microg pUC18 transfected -2



Statistics for: PUC182 , FL2-H v FL1-H , Ungated

Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	5.33
	(X)				4.47
R1	(Y)	0	0.00	0.00	--
	(X)				--

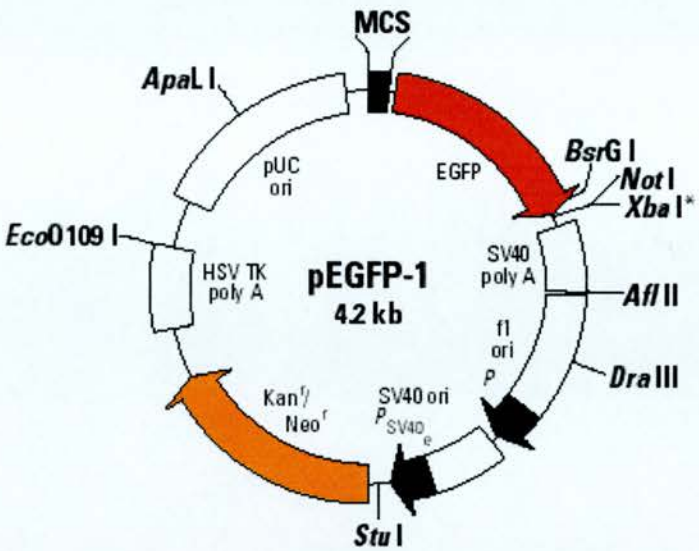
(u) 1 microg pUC18 transfected -3



Statistics for: PUC183 , FL2-H v FL1-H , Ungated

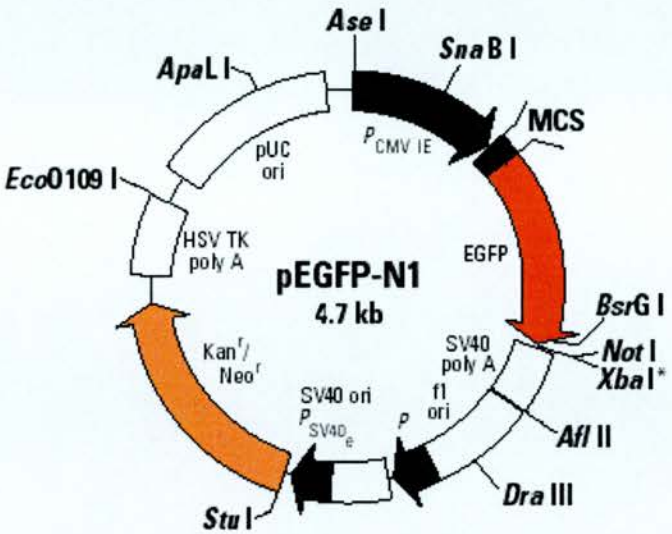
Region		Count	% Gated	% Total	Mean
Total	(Y)	10000	100.00	100.00	4.94
	(X)				4.17
R1	(Y)	0	0.00	0.00	--
	(X)				--

Appendix D: Vector Structure. This appendix shows diagrams of the parent/control vectors pEGFP-1. pEGFP-N, pIRES2EGFP, pCMVβ (figures adapted from the Clontech website) $_{PAC}$RC1b and $_{PAC}$RC2b, along with a brief description of the features in these vectors.



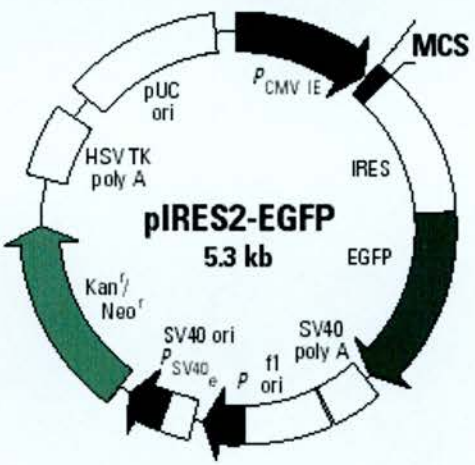
Description

pEGFP-1 encodes a red-shifted variant of wild-type green fluorescent protein GFP which has been optimized for brighter fluorescence and higher expression in mammalian cells. (Excitation maximum = 488 nm; emission maximum = 507 nm.) pEGFP-1 encodes the GFPmut1 variant which contains the double-amino-acid substitution of Phe-64 to Leu and Ser-65 to Thr. The coding sequence of the EGFP gene contains more than 190 silent base changes which correspond to human codon-usage preferences. Sequences flanking EGFP have been converted to a Kozak consensus translation initiation site to further increase the translation efficiency in eukaryotic cells. pEGFP-1 is a promoterless EGFP vector which can be used to monitor transcription from different promoters and promoter/enhancer combinations inserted into the MCS located upstream of the EGFP coding sequence. SV40 polyadenylation signals downstream of the EGFP gene direct proper processing of the 3' end of the EGFP mRNA. The vector backbone also contains an SV40 origin for replication in mammalian cells expressing the SV40 T antigen. A neomycin-resistance cassette (Neo^r) allows stably transfected eukaryotic cells to be selected using G418. The Neo^r cassette consists of the SV40 early promoter, the neomycin/kanamycin resistance gene of Tn5, and polyadenylation signals from the Herpes simplex virus thymidine kinase (HSV TK) gene. A bacterial promoter upstream of this cassette confers kanamycin resistance in *E. coli*. The pEGFP-1 backbone also provides a pUC origin of replication for propagation in *E. coli* and an f1 origin for single-stranded DNA production.



Description

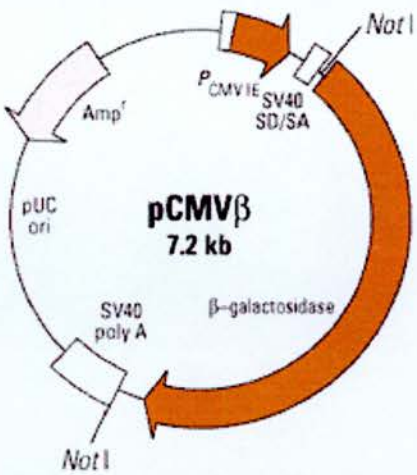
pEGFP-N1 encodes a the EGFP reporter gene. The MCS in pEGFP-N1 is between the immediate early promoter of CMV ($P_{CMV\ IE}$) and the EGFP coding sequences. Genes cloned into the MCS will be expressed as fusions to the N-terminus of EGFP if they are in the same reading frame as EGFP and there are no intervening stop codons. SV40 polyadenylation signals downstream of the EGFP gene direct proper processing of the 3' end of the EGFP mRNA. The vector backbone also contains an SV40 origin for replication in mammalian cells expressing the SV40 T-antigen. A neomycin-resistance cassette (neo^r), consisting of the SV40 early promoter, the neomycin/kanamycin resistance gene of Tn5, and polyadenylation signals from the Herpes simplex thymidine kinase gene, allows stably transfected eukaryotic cells to be selected using G418. A bacterial promoter upstream of this cassette (P_{amp}) expresses kanamycin resistance in *E. coli*. The pEGFP-N1 backbone also provides a pUC19 origin of replication for propagation in *E. coli* and an f1 origin for single-stranded DNA production.



Description

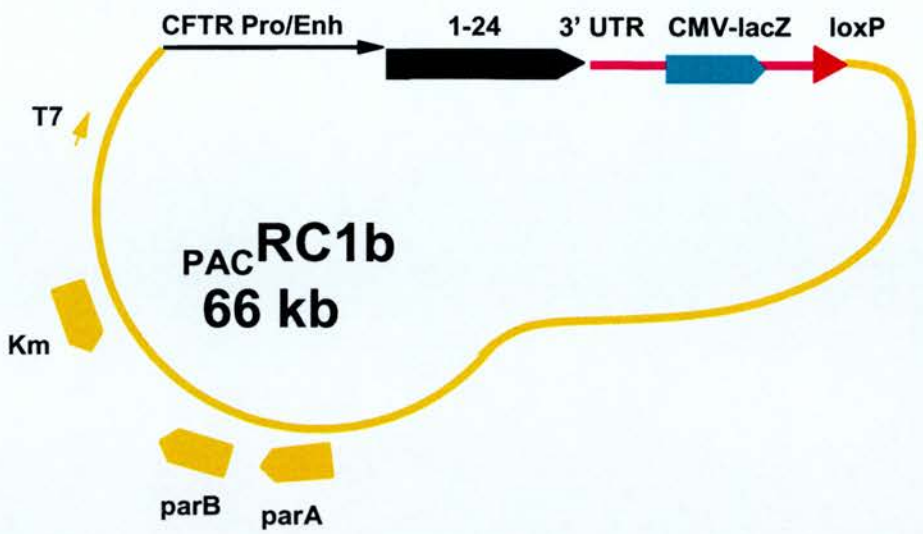
pIRES2-EGFP contains the internal ribosome entry site (IRES) of the encephalomyocarditis virus (ECMV) between the MCS and the enhanced green fluorescent protein (EGFP) coding region. This permits both the gene of interest (cloned into the MCS) and the EGFP gene to be translated from a single bicistronic mRNA. pIRES2-EGFP is designed for the efficient selection (by flow cytometry or other methods) of transiently transfected mammalian cells expressing EGFP and the protein of interest. This vector can also be used to express EGFP alone or to obtain stably transfected cell lines without time-consuming drug and clonal selection.

The MCS in pIRES2-EGFP is between the immediate early promoter of cytomegalovirus ($P_{CMV\ IE}$) and the IRES sequence. SV40 polyadenylation signals downstream of the EGFP gene direct proper processing of the 3' end of the bicistronic mRNA. The vector backbone also contains an SV40 origin for replication in mammalian cells expressing the SV40 T antigen. A neomycin-resistance cassette (Neo^r), consisting of the SV40 early promoter, the neomycin/kanamycin resistance gene of Tn5, and polyadenylation signals from the herpes simplex virus thymidine kinase (HSV TK) gene, allows stably transfected eukaryotic cells to be selected using G418. A bacterial promoter upstream of this cassette expresses kanamycin resistance in *E. coli*. The pIRES2-EGFP backbone also provides a pUC origin of replication for propagation in *E. coli* and an f1 origin for single-stranded DNA production.



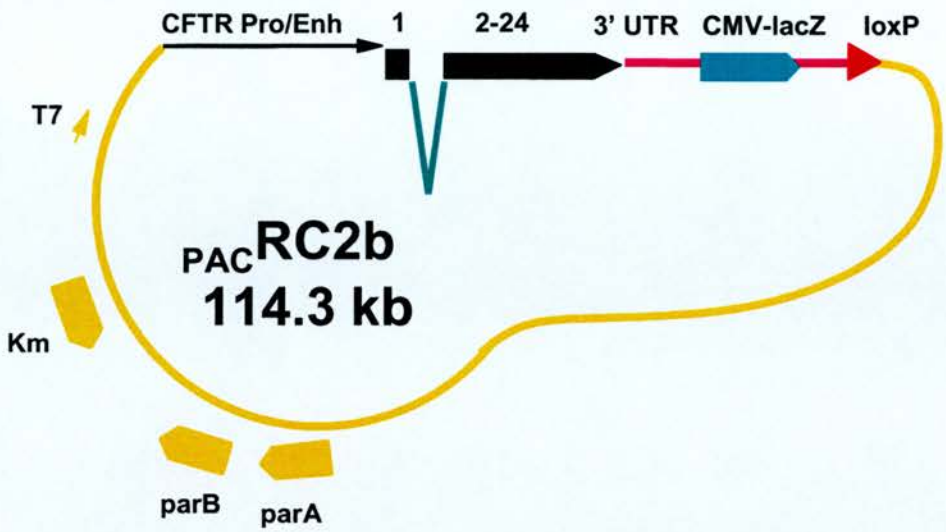
Description

pCMVbeta is a mammalian reporter vector designed to expression beta-galactosidase in mammalian cells from the human cytomegalovirus immediate early gene promoter. pCMVbeta contains an intron (splice donor/splice acceptor) and polyadenylation signal from SV40, and the full-length *E. coli* beta-galactosidase gene with eukaryotic translation initiation signals. pCMVbeta expresses high levels of beta-galactosidase and can be used as a reference (control) plasmid when transfecting other reporter gene constructs and can be used to optimize transfection protocols by employing standard assays or stains to assay beta-galactosidase activity. Alternatively, the beta-galactosidase gene can be excised using the *Not* I sites at each end to allow other genes to be inserted into the pCMVbeta vector backbone for expression in mammalian cells or to insert the beta-galactosidase fragment into another expression vector.



Description

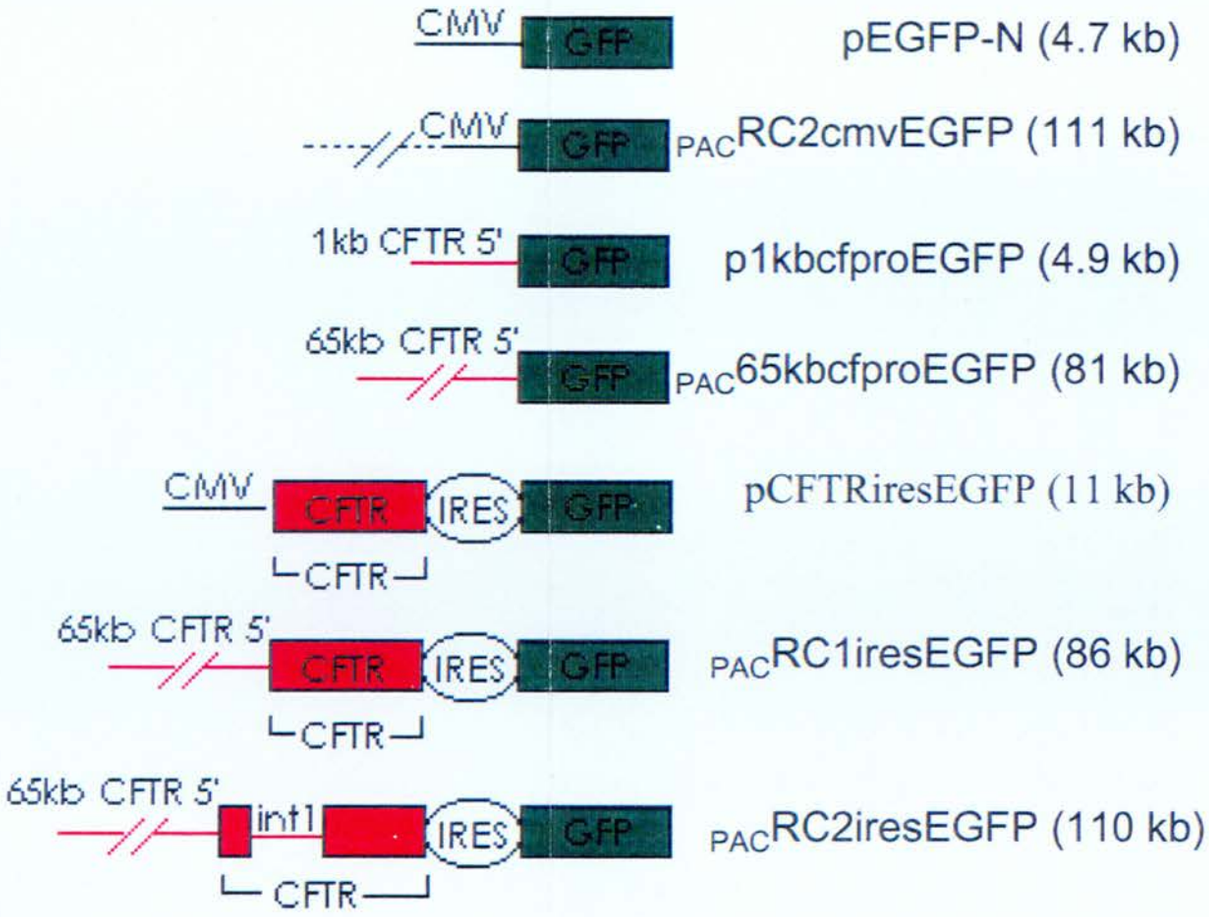
$_{PAC}$RC1b encodes CFTR, including the genomic sequence of CFTR from 65 kb 5' to the ATG start site, then the cDNA sequence from exon 1 through exon 24. Downstream of this, a P^{CMV} promoter drives expression of the βgalactosidase gene. A loxP site allows the vector to be manipulated by a double recombination method (Boyd et al, 1999b). A Kanamycin resistance cassette (Km) allows selection for cells bearing the plasmid in *E. coli.* The vector contains a PAC backbone.



Description

$_{PAC}$RC2b encodes CFTR, including the genomic sequence of CFTR from 65 kb 5' to the ATG start site through intron 1, and then the cDNA sequence from exon 2 through exon 24. Downstream of this, a P^{CMV} promoter drives expression of the βgalactosidase gene. A loxP site allows the vector to be manipulated by a double recombination method (Boyd et al, 1999b). A Kanamycin resistance cassette (Km) allows selection for cells bearing the plasmid in *E. coli*. The vector contains a PAC backbone.

Appendix E: Fold-out diagram of EGFP reporter vectors.



Appendix F: Statistical Analysis.

F1: T-Test

Figure 4.4 (RT-PCR)

"population 1" vs. untransfected

T = 1.269
p=0.2124
Not significant

"population 2" vs. untransfected

T = 4.950
p = 0.0001
Significant

"population 3" vs. untransfected

T = 5275.898
p = 0.0385
Significant

"Population 3" vs. "Population 2"

T = 276.336
p = 0.0023
Significant

Figure 4.9

(a) % cells transfected

Saint vs. DOTAP
T = 40.63
p =0.0157
Significant

Saint vs PEI
T = 23.95
p =0.0266
Significant

DOTAP vs. PEI
T = 4.44
p = 0.1412
Not significant

(b) Xf

Saint vs. DOTAP
T = -2.86
p = 0.2138
Not significant

Saint vs. PEI
T = 9.32
p = 0.0681
Not significant

DOTAP vs. PEI
T = 6.44
p = 0.098
Not significant

Figure 4.10

(a) % cells transfected

Saint vs. PEI
T = 3.51
p = 0.1766
Not significant

Saint vs. DOTAP
T = 8.63
p = 0.0734
Not significant

PEI vs. DOTAP
T = 5.39
p = 0.1167
Not significant

(b) Xf

PEI vs. Saint
T =24.31
p = 0.0262
Significant

PEI vs. DOTAP
T = 35.91
p = 0.0177
Significant

Saint vs. DOTAP
T = 1.90
p = 0.3082
Not significant

Figure 4.14

(a) % cells transfected

SID vs. lipofectin
T = 574.33
p = 0.0011
Highly significant

Saint-mix vs. lipofectin
T = 57.47
p = 0.0111
Significant

LID vs. lipofectin
T = 8.50
p = 0.0745
Not significant

Figure 4.15

(a) % cells transfected

LID vs. lipofectin
T = 38.89
p =0.0082
Highly significant

(b) Xf

SID vs. Saint Mix
T = 9.77
p = 0.0325
Significant

LID vs. Saint
T = 64.75
p =0.0049
Highly significant

SID vs. lipofectin
T = 10.98
p = 0.0289
Significant

LID vs. lipofectin
T =318.98
p =0.001
Highly significant

Saint Mix vs. lipofectin
T =5.75
p = 0.0548
Not significant

Figure 4.16

(a)% cells transfected

LID vs. SID
T = 12.11
p = 0.0445
Significant

LID vs. Saint-Mix
T = 9.60
p =0.033
Significant

LID vs. lipofectin
T =85.58
p = 0.0037
Highly significant

SID vs. lipofectin
T =40.88
p =0.0078
Highly significant

Saint-Mix vs. lipofectin
T =12.12
p =0.0262
Significant

(b) Xf

SID vs. LID
T =7.16

p = 0.0441
Significant

SID vs. lipofectin
T =7.55
p =0.0413
Significant

Saint-Mix vs. lipofectin
T =2.35
p = 0.128
Not significant

LID vs. lipofectin
T =4.94
p = 0.0636
Not significant

Figure 4.17

(a) % cells transfected

LID vs. SID
T =5.39
p = 0.0164
Significant

LID vs. DOTAP
T = 15.28
p = 0.0021
Highly significant

SID vs. DOTAP
T =7.19
p = 0.0094
Highly significant

Figure 4.18

(a) % cells transfected

SID vs. LID
T = 4.72
p = 0.0663
Not significant

SID vs. DOTAP
T =8.65
p =0.0366
Significant

F2. Test statistic $\mu 1\mu 4 = \mu 2\mu 3$

		EGFPN		RC2cmvEGFP		1kbcfproEGFP		65kbcfproEGFP		μ1μ2	μ3μ4
		μ	$\sqrt{\sigma^2}$	μ	$\sqrt{\sigma^2}$	μ	$\sqrt{\sigma^2}$	μ	$\sqrt{\sigma^2}$		
COS-7	Repl 1	785	(257)	481	(197)	714	(135)	423	(127)	1.6	1.7
	Repl 2	800	(253)	506	(262)	704	(166)	413	(137)		
	Repl 3	806	(251)	508	(206)			413	(126)		
	OVER	797	600	499	267	704	421	416	80		
MDCK-IOWA	Repl 1	615	(280)	525	(290)	494	(193)	466	(181)	1.2	1.1
	Repl 2	595	(291)	472	(275)	499	(190)	511	(173)		
	Repl 3	595	(284)	491	(290)	497	(188)	421	(173)		
	OVER	601	388	496	469	496	62	466	316		
T84	Repl 1	517	(223)	432	(182)	474	(177)	442	(159)	1.1	1.1
	Repl 2	487	(220)	441	(207)	452	(178)	415	(214)		
	Repl 3	455	(220)	409	(197)	466	(175)	442	(188)		
	OVER	486	1032	427	93	464	247	432	64		
Caco-2	Repl 1	558	(229)	444	(213)	461	(194)	346	(105)	1.2	1.2
	Repl 2	534	(231)	521	(216)	455	(166)	376	(158)		
	Repl 3	564	(239)	418	(211)	429	(174)				
	OVER	552	394	461	326	408	233	361	122		

Where μ = the mean of the average fluorescence intensity
$\sqrt{\sigma^2}$ = the square root of the mean square

For COS-7:
T = 6.03
Significant difference

For MDCK-IOWA:
T = -0.62
Significant difference

For T84:
T = -1.08
Significant difference

For Caco-2:
T = 0.79
Significant difference